Multimodal travel groups and preference mode for different travel purpose
a latent class cluster analysis

Xiaochen Ma¹, Oded Cats¹, and Serge Hoogendoorn¹

1 Introduction

Mode choice as a part of the travel demand models is essential for assessing policies designed
to influence modal split. Most commonly, these policies involve attracting more people to
reduce car use. It is important to understand the travelers’ behavior to develop those policies.
And from the perspective of scientific, we are able to understand the traveler’s modes use
patterns behaviour, modes use patterns in our research represent travellers’ modes use
frequency and individuals’ travel distance, travel number in a period. Our research can be
helpful to identify (potential) multimodal travellers, that is, travellers who make use of more
than one mode of transport, and not exclusively use a single mode irrespective of context
during a period. (e.g. Aarts et al., 1998). Kroesen (2014) found that multimodal users are
more likely than single-mode users to be encourages to switch modes use frequency
patterns. E.g. They are easily effected by policies to use other transport modes, rather than
single-car users who always use car in every context. Hence, it is indeed necessary to identify
multimodal traveller groups and understand the nature of the group in order to facilitate more
such behaviour. Nobis (2007) found that multimodality is particularly high among
adolescents, older people, and residents of population centres. Blumenberg and Pierce (2014)
found that lower-income Americans are less multimodal than those with higher incomes.
Various researchers have examined that travelers who habitually use the same mode may
have different attitudes and perceptions on other modes. (See, Van Exel and Rietveld (2009).
In specific, for example, that frequent car users who solely use car perceive a longer travel
time towards PT than frequent car users who also travel by PT. (Diana and Mokhtarian,
2009a, 2009b).

To help facilitate their behaviour, e.g, sustainable travel modes uses, it is difficult to
understand the nurture of every individual, because every individual can have their own travel
pattern. The easy way is to group travellers into different travel groups. It is easy and
important to identify and distinguish difference of transport modes use patterns among
travellers’ groups. The approach in most previous analysis is traditional cluster analysis. E.g.
Diana and Mokhtarian (2009a, 2009b), who applied cluster analysis to identify multimodal
travel groups. The traditional cluster analysis, for example, K-means mainly finds clusters
with some arbitrary chosen distance measure. And researchers need to decide the number of
cluster in advance, which easily result in incorrect results. However, latent class analysis can
first describe distribution of the data (indicators) and the latent class is generated based on the
indicators distribution. Second, we assess probabilities that certain cases are members of
certain latent classes. Particularly, in identifying potential multimodal travellers, we can
calculate the probability of the travellers belonging to every group, and then we can assign
travellers into the groups with the highest probability. Latent class analysis should to be
distinguished with latent class choice model, which integrated latent class into choice model
for taking into account to attributes preference heterogeneity of alternatives between groups.
g. Vij et al. (2013) estimated a latent class choice model, the results show different
multimodal style are related to long-term travel decisions and travel time sensitivity.

In transport research domain, latent class analysis could be used to identify unobserved
target-groups with different travel pattern. In particular, in the text of identifying multimodal
clusters by applying LCCA, Diana and Mokhtarian (2009a) investigate the patterns of various

¹Delft University of Technology
multimodal clusters, and they only focused on the socio-demographic variables. Molin and Mokhtaria (2016) identified (multi) modal groups by applying latent class analysis in which the indicators are mode use frequency. Mode perception and attitudes were included as covariates to predict who belongs to multimodal travel groups. Similar with the research on Molin(2016), besides various modes use frequency, in our research we will use travel numbers and distance in one day as indicators to group people.

To the best of the authors’ knowledge, however, mode preferences of travellers have received little attention in the context of identifying multimodal travel groups, partly because it is hard to obtain the data of the self-reported preference required for performing such an analysis. An interesting question is whether travellers’ travel pattern is in line with their preference. For example, some travellers stated their preference is car for different kinds of trips, however they seldom use car, we can say their preference are not in line with their actual behaviour. Based on our results, we can also make conclusion that what are kinds of travellers groups are not consistent between their preference and actual travel behaviour. Because we group people in different classes, therefore in our paper the question that which group has the greatest misalignment with their preference and what are the reasons resulting in such an inconsistency can also be explored. That is, in specific, we can compare travellers’ different transport modes use frequency between their corresponding modes preference in different class, and to analyse their inconsistency reasons according to the traveller’s characteristics and travel pattern. In our research, the meaning that we will take the travel distance and travel numbers per day as indicators as well, is because it can help us understand the travellers in detail. For example, some travellers whose average travel distance is 100km per day (far distance between home and work palaces) stated bike as their preference, so he/she can not use bike everyday. We can derive one of the reasons of inconsistent between their preference and actual travel behaviour is because their actual travel distance requirement. On the other hand, we can group people into detail groups, for examples, people with same modes frequency patterns can be grouped into different behaviour class, because of their different travel distance per day. Then we can understand and guide various groups’ people behaviour in different specific ways.

In conclusions: our researches will contributes in the art of identifying travelers travel patterns groups. One contribution is the indicators we use for clustering include mode use frequency and individual travel patterns, which are average travel number and average trip distance per day. Second contribution is, based on our results, we can also make conclusion that what are kinds of travellers groups are not consistent between their preference and actual travel behaviour, and derive the corresponding reasons. Our research will provide various suggestions for transport policy, for example how to facility the travellers’ preference modes uses.

The paper is organized as follows: in section 2, we summarize the methodology and theoretical formulation for latent class analysis. Section 3, we describe the data process and analysis and implementation. Section 4, we describe and interpreter the results. Finally, section 5, we conclude the paper and identify possible further studies.

2 Methodologies

2.1 Latent class analysis theory

In this paper, travellers will be classified based on their reported mode use frequency, average travel distance and average travel numbers per day by using latent class analysis (LCCA). The approach for distinguishing the different travel groups is similar to the approach taken by Molin (2016). The main idea of LCCA is that a discrete latent variable can account for the observed associations between a set of indicators, such that, conditional on the latent class variable, these associations become insignificant (Magidson and Vermunt, 2004; McCutcheon, 1987). This is generally called the assumption of local independence. An advantage of the LCCA over conventional cluster analysis is that the approach assigns travellers to clusters probabilistically. Latent class analysis starts with describing distribution
of indicators variables, while other clustering algorithms start with finding similarities between cases. The approach enables researchers to assign people into different groups. This reduces misclassification biases.

Latent class cluster analysis includes covariates to predict the likelihood of individual belonging to which latent class membership. In general, this analysis includes two steps: First, identify groups of the travellers according to the frequency of use of various transport modes in a period and average travel distance and numbers per day. Second is a membership model, which predicts the probability of individuals belonging to each of the identified clusters, based on the indicators. There are also distributions in different clusters of geographical profiles, and preference mode for different purpose trips in our research.

2.2 Latent class analysis model

In latent class model the latent variables $x$ is being measured by the observed variables or indicators. Adapted from Magidson(2004), We will present the model by using formulas below:

$$\pi^{abx}_{ijt} = \pi^{x}_{ti} \pi^{ai}_{it} \pi^{bx}_{jt}, i = 1,...,I; j = 1,...,J, t = 1,...,T$$  \hspace{1cm} (1)

In the formula (1), $a$ is an observed variable having $I$ classes($i = 1,2,...,I$), variable $b$ is an observed variable having $J$ classes($j = 1,2,...,J$), and variable $x$ is an unobserved or latent variable having $T$ classes, $\pi^{abx}_{ijt}$ denote the joint probability that an observation is in class $i$ on variable $a$, in class $j$ on variable $b$, and in class $t$ on variable $x$; let $\pi^{x}_{ti}$ denote the conditional probability that an observation is in class $i$ on variable $a$, given that the observation is in class $t$ on variable $x$ and let $\pi^{bx}_{jt}$ denote the conditional probability that an observation is in class $j$ on variable $b$, given that the observation is in class $t$ on variable $x$, and let $\pi^{abx}_{ijt}$ denote the probability that an observation is in the class $i$ on variable $A$ and in class $j$ on variable $B$, given that the observation is in class $t$ on variable $X$. this model states the variables $A$ and $B$ are conditionally independent of each other, given the class level on variable $X$, that is,

$$\pi^{abx}_{ijt} = \pi^{x}_{ti} / \pi^{a}_{i} = \pi^{bx}_{jt} / \pi^{b}_{j} \pi^{abx}_{ijt}$$  \hspace{1cm} (2)

Formula (2) is the conditional probability that an observation is in in class $I$ on variable $A$ and in class $J$ on variable $B$, given that the observation is in class $T$ on variable $X$.

3 Data and Estimation

3.1 Data selection

For our analysis, we use data from the Netherlands Mobility Panel (MPN) that were collected in the period 2013-2015 (Hoogendoorn-Lanser, 2015). This representative sample of the Dutch population comprises over 2500 households surveyed in three (on-year) waves. The MPN is a household panel that has been set-up in order to study the short-run and long run dynamics in travel behaviour of Dutch individuals and households, and to determine how changes in personal and household characteristics and in other travel-related factors correlate with changes in travel behaviour. It consists of a screening questionnaire and a household questionnaire that are filled out by an adult household member (gatekeeper), and of two individual questionnaires and a travel diary that are filled out by each household member of
twelve years or older. Annually, respondents are asked to fill in these questionnaires and the travel diary. For our analysis, we use data from the questionnaires.

A sample from all panel members older than 12 years was drawn from this panel; those who were invited to take the survey were selected at random. The dataset consisting of three days’ travel diary surveys of 2500 respondents (older than 12 years old) in 2014. The respondents reported their frequency of using various modes, which varies from more than 4 days in a week to less than once in a year. Which are ordinal variables that from 1 to 6, representing the frequency of various modes use. Respondents’ self-reported frequency of modes in one year though has less-detail than diary data, e.g., Nobis (2007), Buehler and Hamre (2015) and Kuhnlimhof et al (2012) used one-week travel diary data, and Vij et al (2013) relied on six-week travel diary data. Self-reported frequency in one year represents travellers ‘mode use patterns better than travel diary, which are recorded in short time. Since multimodality is defined as use of various transport modes in a certain time period (Nobis, 2007). So it is better to classify a traveller as a (non) multimodality based on a longer time period than short time. The subsample on which the analyses in this paper are based on totals 618 respondents. The data finally used for latent class analysis contains information about socio-demographic characteristics of the respondents, their self-reported frequency of modes use, which contain car, urban public transport, train and bike. And preferred modes for work and non-work trips.

3.2 Model estimation

In our research, the indicators we use for clustering travellers are modes frequency in a period, trip numbers and travel distance per day. Specifically, the four main modes are car, train, BTM (bus, tram, metro) bike and walk. Respondents were asked to report how often they use each of the four distinguished modes on a six-point ordinal scale ranging from (1) more than 4 days every week to (6) less than once a year. And the other two indicators are average travel distance and times per day. The two variables are calculated by using the 3-days travel diary in our data. Specifically, we calculated the average travel distance and travel numbers recorded in the 3-days diaries. Travel number here means the number people travel on one-day, including all kinds of traveling.

As shown in Fig.1, a range of socio-demographic variables (except car number) will be explored as covariates in our model to explain the membership of each of the identified traveller clusters. Perceptions of individual on their residential regions and the population density are showed in Fig. 2.
There are 3 main approaches to applying latent class analysis in doing research, Haughton, et al (2012), namely, Latent GOLD, MCLUST and poLCA, in our research, poLCA is used. We will present the results of the latent class analysis in next section.

4. Results

By applying the latent class analysis, the results show that the 618 respondents are classified into classes based on their self-reported modes use frequency and travel distance and numbers, which we group them from 1 to 6 in the table below. There is Bayesian information criterion (BIC) and AIC information criteria in the result. Formulas below (3) and (4), BIC is preferred over AIC in latent class models, which is a criterion for model selection among a finite set of models. A smaller BIC is better than a bigger BIC. Next to AIC and BIC from the result we can also get a Chi-Square goodness of fit. According to statistics, group people into 4 groups are good.

\[
BIC = \ln(n)k - 2\ln(L)
\]
\[
AIC = -2\ln(L) + \ln(n)\times k
\]

In the formulas, \(k\) means the parameters, \(L\) means the loglikelhood function, and \(n\) means the observations number.

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>9016</td>
<td>8610</td>
<td>8492</td>
<td>8423</td>
<td>8411</td>
<td>9240</td>
</tr>
<tr>
<td>BIC</td>
<td>9126</td>
<td>8898</td>
<td>8879</td>
<td>8838</td>
<td>9071</td>
<td>10037</td>
</tr>
<tr>
<td>Loglikelihood</td>
<td>-4483</td>
<td>-4249</td>
<td>-4159</td>
<td>-4093</td>
<td>-4056</td>
<td>-4440</td>
</tr>
<tr>
<td>Chi-square goodness of fit</td>
<td>38427</td>
<td>12554</td>
<td>12355</td>
<td>12194</td>
<td>12297</td>
<td>38424</td>
</tr>
</tbody>
</table>

Table 1. Model fit of the latent class cluster analysis models.

Figure 2 Perceptions of individual on their residential regions and the population density
4.1 Within-cluster distribution of indicators

Because the indicators of trip numbers are almost same between groups, we will not show the indicators then. In Table 2, we describe the distributions of 5 indicators responds in four clusters. The indicators’ names are on the left columns. Based on the distribution of five indicators we classify the 618 respondents into four groups. The four clusters account for 27%, 19%, 22% and 32% respectively. For the interpretation of the results we mainly rely on the within-cluster distributions of the indicators. We name them as 1) long distance private modes travellers, 2) middle distance public transport and bike travellers, 3) short distance multi-mode travellers, and 4) short distance private modes travellers. Specifically, for example, the first group is defined as long distance private modes travellers group, because people who have higher probability of using private modes (car and bike) and higher probability of having average long distance trips per day, have the most highest probability belonging to this group. In detail, if a traveller uses car more than 4 days a week, and use train and BTM 2 days in a year, bike 1-4 days a week, and have average trip distance 40km per day. The latent class analysis holds the assumptions that there is local independence in groups, which means the indicators’ responds in every group are independent. Hence, the probability of the people belongs to which group is joint probability of the indicators multiply the group proportion. The probability of the individual belonging to first group equal to

\[
\text{94\%} \times \text{74\%} \times \text{50\%} \times \text{44\%} \times \text{12\%} \times \text{27\%} = 99.999\% 
\]

Since, it is based on the joint probability of giving the respondents on the indicators, if there is some missing data value of the indicators responds, we can also calculate the probability. Similarly, when the responds of the five indicators are given, we can calculate the probability of the people belonging to every group. Then we assign travellers to the group with the highest probability in the membership model. In next part, we will introduce the results of the membership model in our research.

Table 2. Within-cluster distributions of the indicators

<table>
<thead>
<tr>
<th>Class population</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.More than 4 days a week</td>
<td>27%</td>
<td>19%</td>
<td>22%</td>
<td>32%</td>
</tr>
<tr>
<td>2.1-4 days a week</td>
<td>94%</td>
<td>16%</td>
<td>4%</td>
<td>61%</td>
</tr>
<tr>
<td>3.6-11 days a month</td>
<td>0%</td>
<td>19%</td>
<td>13%</td>
<td>6%</td>
</tr>
<tr>
<td>4.1-3 days a month</td>
<td>0%</td>
<td>13%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>5.1-5 days a year</td>
<td>0%</td>
<td>8%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>6.less than 1 day a year</td>
<td>0%</td>
<td>6%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>Train frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.More than 4 days a week</td>
<td>0%</td>
<td>17%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>2.1-4 days a week</td>
<td>0%</td>
<td>15%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>3.6-11 days a month</td>
<td>0%</td>
<td>45%</td>
<td>7%</td>
<td>0%</td>
</tr>
<tr>
<td>4.1-3 days a month</td>
<td>15%</td>
<td>15%</td>
<td>29%</td>
<td>0%</td>
</tr>
<tr>
<td>5.1-5 days a year</td>
<td>74%</td>
<td>6%</td>
<td>36%</td>
<td>0%</td>
</tr>
<tr>
<td>Frequency</td>
<td>More than 4 days a week</td>
<td>1-4 days a week</td>
<td>3-6 days a month</td>
<td>6 days a month</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Bike frequency</td>
<td>2%</td>
<td>16%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Trip distance</td>
<td>3%</td>
<td>15%</td>
<td>13%</td>
<td>14%</td>
</tr>
</tbody>
</table>

### 4.2 Membership model

Based on our analysis above, we group the 618 individuals into four groups based on five indicators. In the membership model, the individual are assigned into the group which have the highest probability. Finally, the predicted class memberships (by modal posterior prob.) are 27%, 18%, 20%, 35% respectively. Call back, The membership share are almost same with estimated class population shares above, which are 27%, 19%, 22%, 32%.

1) The within-cluster covariates and car number distributions

In the figures (1A-4A) below, we show the 5 characteristics (covariates) of the respondents and the variable of car number of family in the four clusters. **Fig.5** shows the Population Density of Four Classes and Sample. In detail, Very highly urbanized: 2500 or more inhabitants/km², Highly urbanized: 1500 to 2500 inhabitants/km², Moderately urbanized: 1000 to 1500 inhabitants/km², Low urbanization: 500 to 1000 inhabitants/km², Non-urbanized area: Less than 500 inhabitants/km².

**Fig.6** shows perceptions of the accessibility on car, bike, btm, park of the four classes and sample. Specifically, people were asked to answer their perceptions on their residential regions accessibility to car, bike, btm, park. Questions are, my neighborhood has a sufficient number of parking places, My neighbourhood has a good cycling infrastructure, my neighbourhood is easily accessible by Public Transport, and my neighborhood is easily accessible by car.
Fig. 1A: Covariates and Car number variable of First Class

Fig. 2A: Covariates and Car number variable of Second Class

Fig. 3A: Covariates and Car number variable of Third Class
**Fig. 4A**: Covariates and Car number variable of Fourth Class

**Fig. 5**: Population Density of Four Classes and Sample
Fig. 6: Perceptions of the accessibility of the four classes and sample

1) **Long distance private modes travellers:** Individual who have the mobility patterns of long distance private modes use, are predominantly male, with compared lowest income and education. Compared to other groups, they have more children and are younger. Just 1% of them do not have a car in their home. Most of them live in high urbanization areas, however, we need to note, there are almost 10% of them live in very low urbanization areas.

2) **Middle distance public transport and bike travellers:** Individual who have the mobility patterns of Middle distance public transport and bike use are predominantly female, with compared higher income and lower education. Compared to other groups, they have less children and middle aged. 18% of them do not have car in their home. Most of them live in high urbanization areas, seldom live in very low urbanization areas.

3) **Short distance multi-mode travellers:** Individual who have the mobility patterns of Short distance multi-mode use have similar gender probability, with compared higher income and highest education. Compared to other groups, they have less children and younger. 45% of them do not have car in their home. Most of them live in very high and high urbanization areas.

4) **Short distance private modes travellers:** Individual who have the mobility patterns of short distance private modes use are slight more female, with compared highest income and higher education. Compared to other groups, they have more children and younger. Just 5% of them do not have car in their home. Most of them live in moderately urbanization areas, lower areas. 10% of them live in very lower urbanization areas.
In conclusion, there are 3 main findings of the memberships. First, we find people who have the travel patterns of using private transport modes most, have more children and live in lower urbanization areas. Many of them almost never use public transport, especially urban public transport (bus, tram, metro). And more than 95% of them own at least one car at home. Most of them are younger, and there is not same disturbance of income and education of people who have the travel pattern of using private transport modes most. Second, individual who use public transport and multi-modes are female and have higher income. Most of them live in very high and high urbanization areas. They have less children and at least 18% of them do not have a car. Especially, the multi-modes travellers, 45% of them do not own a car. People who belong to these groups travel less distance than other groups ‘people. Third, according to the figure6A, we find individuals’ perceptions on car, bike, BTM, park accessibility are in line with their modes use frequency patterns. For example, people in group2 and group 3, have stronger positive on BTM and bike accessibility.

4.3 The within-cluster covariates

As mentioned above, we include five covariates in our latent class analysis. The five covariates are socio-demographic variables include gender, age, education, family income, children number. We assume the five sociodemographic variables can effect individuals’ travel behaviour. In our research, we assume these five variables effect the travelers’ travel patterns. The within-cluster distributions of the covariates as presented in Table 3 below. Comparing to group 1 and group 2, based on the T-values, there are three covariates effect the probability of individuals’ belonging to which group. They are education, income and children number. The coefficient are 0.40, -0.23 and -0.89. Which means the higher education the individual have, the higher probability they belong to class 2 rather than class 1. And the negative coefficients of significant variables of income have children number mean that, the less income and children the people have the higher probability they belong to class 2, compared with class 1. Which are all in line with the distribution of socio-demographic variables. In table 3 (2/1) means, compared to class 1, the compared probability of people belonging to class 2. Similar, compared with class 1, male and people who have lower education, older age have higher probability belong to class 3. The negative intercept -3.21 means comparing to class 1, individual have a constant lower probability belong to class 3. This is in line with the proportion of the class 1 and class 3, 27% and 22%. Comparing to class 1, people with higher education have lower probability belong to class 4, which also are consistent with the distribution of socio-demographic variables in classes.
Table 3. The within-cluster distributions of the covariates

|       | Coefficient | Std.error | T value | Pr (>|t|) |
|-------|-------------|-----------|---------|---------|
| 2/1   |             |           |         |         |
| (Intercept) | -1.76      | 1.56      | -1.13   | 0.26    |
| Children_N | -0.89      | 0.36      | -2.46   | 0.01    |
| Education | 0.40       | 0.12      | 3.29    | 0.00    |
| Income  | -0.23      | 0.11      | -2.19   | 0.03    |
| Gender  | 0.47       | 0.32      | 1.44    | 0.15    |
| Age     | -0.05      | 0.08      | -0.61   | 0.54    |
| 3/1   |             |           |         |         |
| (Intercept) | -3.21      | 1.58      | -2.04   | 0.04    |
| Children_N | -0.16      | 0.37      | -0.42   | 0.67    |
| Education | -0.22      | 0.11      | -2.2    | 0.03    |
| Income  | -0.11      | 0.11      | -1.00   | 0.315   |
| Gender  | 0.77       | 0.34      | 2.30    | 0.02    |
| Age     | 0.27       | 0.08      | 3.33    | 0.00    |
| 4/1   |             |           |         |         |
| (Intercept) | 1.38       | 1.29      | 1.07    | 0.29    |
| Children_N | 0.11       | 0.18      | 0.60    | 0.55    |
| Education | -0.23      | 0.10      | -2.21   | 0.03    |
| Income  | -0.14      | 0.09      | -1.53   | 0.13    |
| Gender  | 0.31       | 0.27      | 1.15    | 0.25    |
| Age     | 0.00       | 0.07      | 0.11    | 0.91    |

4.4 Preference transport modes for different purpose trips

One of our research goals is to reveal the relation between travelers’ actual transport use frequency patterns and their preference transport modes. Especially, we classify the travelers into four groups based on their self-reported various transport modes use frequency and travel distance per day. We can compare different groups’ transport modes preference and the relation between their modes preference and use frequency. Our results also can give us interpretations of the reasons about the inconsistency between transport modes use frequency and preference. For example, some travellers whose average travel distance is 100km per day (far distance between home and work palaces) stated bike as their preference, so he/she cannot use bike everyday. We can derive one of the reasons of inconsistent between their preference and actual travel behaviour is because their actual travel distance requirement.

Here in the fig7 below, four different purpose trips are defined, which are work, shopping, hang out and one day trip. Five colors mean four groups and sample group in the picture. We can see for four kinds of different purpose trips, the cars are all rank first. Especially for the one day trip trips, average 55% of all the five groups prefer cars as their transport modes. The private modes users of class one and class four all prefer car most than other groups in every kinds of trips. For the class 2 and class 3, they do not prefer car much in every kinds of trips. Which are consistent with their actual travel modes use frequency too.

One interesting finding is for all kinds of purpose trips, people in group one have a very lower preference on bike. The reason we can derive is that this group people have a long
average travel distance per day. In other words, bikes can not meet their requirement of long distance trips. Hence, when distances of the four kinds of trips are longer, they prefer to use car, rather than bikes. Even though they use bikes often, hence, we can devier the cycling trips are shorter. In the Netherlands, bike is popular as the access and egress parts for public transport. Therefore, one reason that people in public transport and multi-modes classes prefer bikes is they usually use bikes for accessing and egressing public transport usually. Train is popular for one day leisure trips. In conclusions, in our research, individuals’ actual behaviour are in line with their modes preferences.

5. Conclusions and limitations

In this paper, latent class cluster analysis was applied to identify groups and to explore different travel patterns and the effects of socio-demographics on the probability of belonging to each of the four groups. Based on the distributions of five indicators, we classify people in four classes. For the interpretation of the results we mainly rely on the within-cluster distributions of the indicators, hence we name them: 1) long distance private modes travellers, 2) middle distance public transport and bike travellers, 3) short distance multi-mode travellers, and 4) short distance private modes travellers. On one hand, When the respondents of the five indicators are given, the probability of the individual belong to every class can be calculated. Since the latent class analysis theory holds the assumption that, there is a latent variable can be divided in to category (class), and in every category, the indicators are local independent. It equals to the joint distribution probability of indicators respondents in every group multiply the proportion of the group and then are divided by the sum of numbers of all the groups. Hence, on the other hand, when we know which group the individual highly belonging to, the joint probability of the indicators respondents can be calculated as joint distribution probability. Which meet the local independent theory of latent class cluster analysis.

In the membership model, we describe the distribution of socio-demographic variables, car-owner variable, and residence regions variables of all the four classes. Our findings are in line with the findings of Nobis (2007), who found that multimodality is particularly high among adolescents, older people, and residents of population centers. And the class 2 and

![Preference modes for different purpose trip](image_url)

**Fig. 7**: Preference transport modes for different purpose trips for four groups.
class 3 groups, in which people have higher probabilities in using public transport and multi-modes in a period have compared lower family incomes.

In the membership model, we include five socio-demographic as covariates. Comparing to group 1 and group 2, we find three covariates effect the probability of individuals’ belonging to which group. They are education, income and children number. Similar, among with class 1 and class 3, gender, age, education and constant intercept are significant effect the probability of belonging to which groups between class 1 and class 3. And education is the only significant covarites that effect the belongings probability between class 1 and class 4. In our research, we also compare actual travel modes use patterns and preference travel modes. We find every groups actual travel patterns are in line with their preference modes. Except people in group one have a very lower preferency on bike, though they have high frequency of using bike. The reason we can derive is that this group people have a long average travel distance per day. In other words, bikes can not meet their requirement of long distance trips. Hence, when distances of the four kinds of trips are longer, they prefer to use car, rather than bikes. Even though they use bikes often, we can derive their cycling trips distances are shorter.

Based on the results presented in this paper we next discuss the challenges for policymakers. For the four groups, we can provide different policy to attract and encourage their sustainable travel behaviour. For example, providing more travel discount tickets for one day trip. Provide convenient shuttle for long distance commuters, because they can reduce their car uses. Policy also should attractive people’s perceptions or preference on sustainable modes. Even though based on our research results, we do not know what are the causal relations between preference and frequency on modes. They are consistent. Hence, policy can try to attractive people more preference on sustainable modes, to examine their influences on modes use frequency. The transport department and government can provide public transport between city center to suburban areas. Since most of public transport are in high density urban areas, the people who live in suburban areas need to drive car to city center, it will result in parking and congestions problems. Government should encourage younger people and low-educated people use sustainable modes, since in group 1, most of such individual usually use car.

Limitation: In our research we scope the covarites variables in socio-demographics variables. However there are other variables also can be seen as covarites, which effect individuals’ travel patterns. And in future, the transition of the latent class can be explored for understanding individual travel behavior changes. Those are important to predict travel behaviors.
Reference:


(10) PoLCA: An R Package for Polytomous Variable Latent Class Analysis


