Descriptive analysis: survey ”Mobility dynamics in urban modes of transportation in Cali, Valle del Cauca, for the definition of public policies and environmental management”.

Project: Analysis of the dynamics of urban mobility by
by motorcycle in the city of Cali for the definition of public
policies and environmental management*

[Working paper 05/23]

Contents

*Pontificia Universidad Javeriana — Cali

List of Figures

List of Tables

1 Introduction

Within the framework of the project "Analysis of the dynamics of urban mobility by motorcycle in the city of Cali for the definition of public policies and environmental management", a survey is conducted with the purpose of reporting information on travel habits in urban transport modes in Cali, Valle del Cauca. Based on recent literature, variables are identified that, in general terms, provide useful information on the factors that influence the choice of urban transport modes. Consider the following variables of interest: **(1)** vehicle purchase and operation cost, **(2)** safety in terms of accidentality, **(3)** safety in terms of crime, **(4)** comfort, **(5)** travel time, **(6)** air pollutant emissions, and **(7)** friends' and relatives' perception of the transport mode (see Fu, 2021; Gadepalli, Tiwari, & Bolia, 2018; S., Vabuolytė, Burinskienė, & Antuchevicienė, 2020; Ahmed, Catchpole, & Edirisinghe, 2020; Agarwal, Ziemke, & Nagel, 2020; Fan & Chen, 2020; Ryan, 2020; Faboya, Ryan, Figueredo, & Siebers, 2020). For description of variables, see the Survey Dictionary.

This is a descriptive cross-sectional study based on the application of 571 voluntary and structured surveys in different places of interest in the city of Cali: transit offices, automotive diagnostic centers, administrative offices, classrooms and common spaces in universities in the city. Additionally, the inclusion criteria consisted of frequent users of some means of urban transportation in the city of Cali. In methodological terms, the selection of each participant in the survey operates by simple random sampling such that the selection of the units is equiprobable. Formally, let N be the population size, and let n be the sample size. Since the sample selection is equiprobable, it is verified that the probability of selecting a sample $s = (u_1, ..., u_n) \in S$, where $|S| = \binom{N}{n})$, is given by

$$\phi(s) = \frac{1}{\binom{N}{n}} \tag{1}$$

Similarly, if the selection of units operates by simple random sampling without replacement, the probability of selecting a first unit u_i, para $i \in \{1, ..., n\}$, is given by

$$\phi(u_i \in s) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N} \tag{2}$$

By hypothesis, the target population corresponds to the working age population in Cali, Valle del Cauca (1,837,000). The sample size calculation is based on a dichotomous variable, namely sex. Following Krejcie and Morgan (1970), since it is considered a dichotomous variable, assume an acceptable margin of error of 5%. Additionally, assume that the probability of committing type I error is also 5%.Cochran (1977) formula with finite population correction establishes that the sample size is given by

$$n = \frac{n_0}{1 + (n_0 - 1)/N} \tag{3}$$

where

$$n_0 = \frac{t^2(p)(1-p)}{d^2} \tag{4}$$

In general, t expresses the t-value for $\alpha = 0.025$ in each tail of the distribution —i.e.: $t = 1.96$—. Likewise, $p(1-p)$ corresponds to the estimated variance; and d, to the acceptable margin of error —i.e.: $d = 0.05$—. By hypothesis, it is determined that $p = 1/2$ in order to obtain the maximum estimated variance and, consequently, the maximum sample size. (Note that, since the selected sample represents less than 5% of the population, the correction factor for finite population of Cochran (1977) is not necessary).

2 Statistical analysis

Note that, for the 571 frequent users of urban transport modes in Valle del Cauca, not only are measurements of the variables of interest **(1) - (7)**; recorded, but also the following economic and socio-demographic characteristics are reported: sex, age, occupation, main activity, socio-economic stratum, city and neighborhood of residence (currently the document is in process of being updated with new responses from users).*Annex 3* presents the data dictionary for the survey, which provides the following characteristics for each variable: name, class, type, description, categories —if applicable— and length. (For the general data structure, see **Table 1** and **Table 2**).

In general terms, in the case of continuous variables, there are two types of results: first, if the continuous variable follows a normal distribution, the mean and standard deviation are presented; second, if the continuous variable is non-Gaussian, the median and interquartile range are presented. The normality test corresponds to the Shapiro-Wilk test. In the case of categorical variables, the count and the proportion are simply presented. Similarly, the methods of comparisons between groups differ according to the type of variable: for continuous variables, the comparison operates according to the Mann-Whitney nonparametric test; for categorical variables, according to the χ^2 test or Fisher's exact test.

3 Results

3.1 Missing value analysis

In the following, an analysis of the number of missing values reported for each variable is presented. Figure 1 represents the missing values differentiated according to the rows of the sample. In general, the figure considers that variables measuring optional responses are omitted in the missing values analysis. Note that the missing information corresponds to optional responses and, therefore, they are kept in the subsequent descriptive analysis.

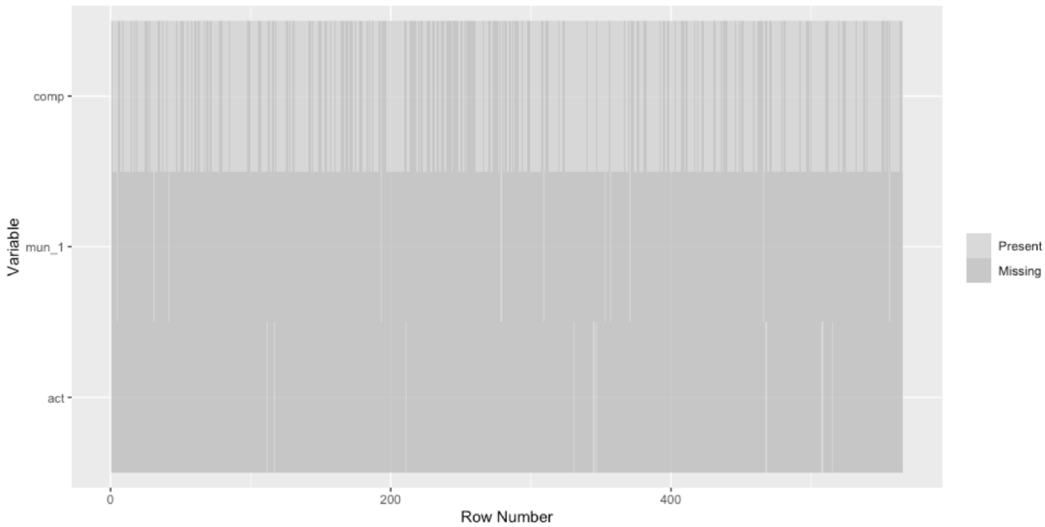


Figure 1: Missing information differentiated according to observations (rows)

3.2 Sociodemographic characterization

In terms of participation, the following cross-tabulations are considered: first, Table 1 presents the proportion of users of urban modes of transport differentiated by sex and age;

Table 2,the proportion differentiated by main mode of urban transport and age; and Table 3, the proportion differentiated by main mode of urban transport and socioeconomic stratum. Note that, in general, the following recoding on socioeconomic stratum is considered: users in low-stratum housing correspond to users in low-low (stratum 1) or low (stratum 2) housing; users in medium-stratum housing correspond to users in medium-low (stratum 3) or medium (stratum 4) housing; and users in high-stratum housing correspond to users in medium-high (stratum 5) or high (stratum 6) housing.

Table 1: Proportion (%) of participants differentiated according to gender and age groups age groups

	Female	Male	Prefer not to answer
15 - 19 years	7.69	8.17	0.32
20- 24 years	9.29	19.23	0.16
25- 29 years	3.21	3.04	0.00
30- 34 years	4.97	4.17	0.00
35- 39 years	4.81	3.69	0.00
40 - 44 years	4.49	4.81	0.00
45- 49 years	4.01	3.53	0.00
50 - 54 years	3.21	2.24	0.00
55 - 59 years	2.08	3.53	0.00
60 -64 years	0.48	2.24	0.00
65 - 69 years	0.00	0.16	0.00
70 or older	0.16	0.32	0.00

Note: Participants correspond to frequent users of urban transport modes.

Table 2: Proportion (%) of participants differentiated by mode of transport and age groups

	App	Car	Motorcycle	Other	Taxi	Informal	Public
15 - 19 years	1.28	7.21	0.80	0.96	0.00	0.48	5.45
20- 24 years	2.56	13.78	3.85	2.08	0.00	0.64	5.77
25- 29 years	0.32	2.40	2.08	0.80	0.00	0.16	0.48
30- 34 years	0.96	4.49	1.92	1.12	0.00	0.00	0.64
35- 39 years	0.16	4.33	2.24	0.80	0.00	0.16	0.80
40 - 44 years	0.96	4.01	1.76	1.44	0.16	0.16	0.80
45- 49 years	0.16	5.45	1.28	0.48	0.00	0.00	0.16
50 - 54 years	0.00	4.49	0.16	0.48	0.00	0.16	0.16
55 - 59 years	0.32	2.88	1.60	0.48	0.00	0.00	0.32
60 -64 years	0.00	2.24	0.32	0.16	0.00	0.00	0.00
65 - 69 years	0.00	0.00	0.00	0.00	0.00	0.00	0.16
70 or older	0.00	0.32	0.00	0.00	0.00	0.00	0.16

*Note: Participants correspond to frequent users of urban transport modes. The category **Other** corresponds to cycling, skateboarding, walking, etc.; **Public transport**, to formal public transport —MIO, e.g.—; and the category **App**, , to applications such as Uber, Didi, Cabify, etc.*

Table 3: Proportion (%) of participants differentiated by mode of transport and socioeconomic stratum

	High	Low	Medium
App	2.88	0.16	3.69
Car	25.64	1.60	24.36
Motorcycle	1.60	4.81	9.62
Other	3.21	0.48	5.13
Taxi	0.16	0.00	0.00
Informal transport	0.16	0.80	0.80
Public transport	1.92	3.53	9.46

*Note: Participants correspond to frequent users of urban transport modes. The category **Other** corresponds to bicycles, skateboards, walking, etc.; **Public transport**, , to formal public transport —MIO, e.g.—; and the category **App**, to applications such as Uber, Didi, Cabify, etc. The**Low** stratum includes strata 1 and 2; the **Medium** stratum, strata 3 and 4; and the **High** stratum, strata 5 and 6.*

3.3 Georeferenced information

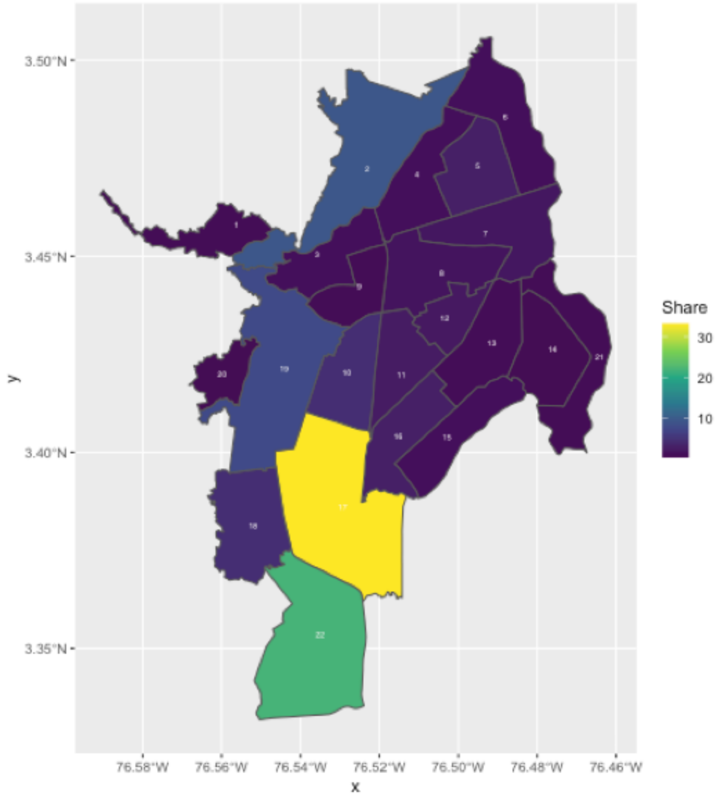
For the case of Cali, the georeferencing of the information is considered in two senses: on the one hand, Figure 2a represents the proportion according to the communes in which the users of urban transport modes reside in the city; Figure 2b, on the other hand, represents the proportion according to the communes to which the users usually travel. Note that, in general, the geo-referencing does not consider **(i)** users who reside in other nearby municipalities and **(ii)** users who commute to a different area.

3.4 Descriptive summary

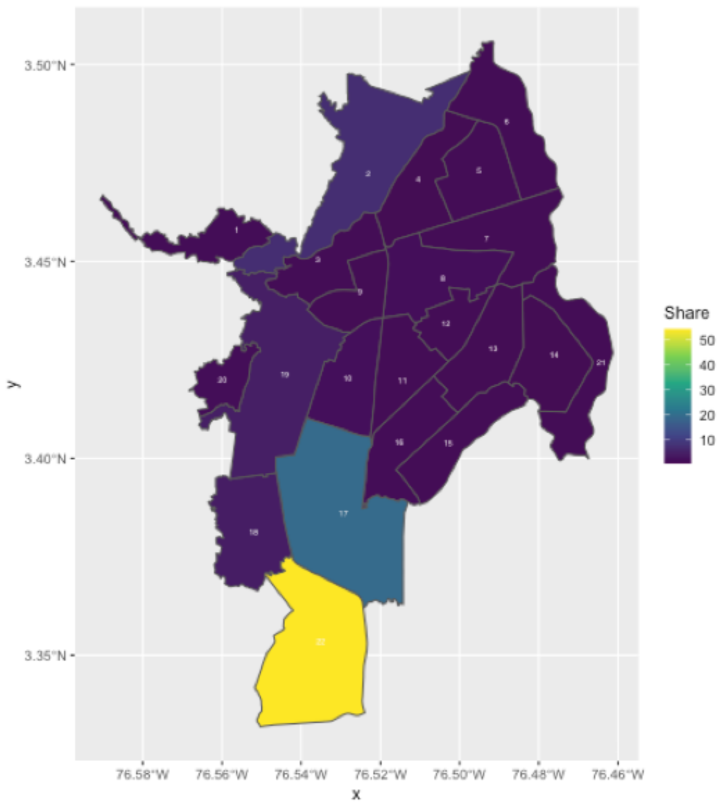
3.4.1 Categorical and continuous variables

In general, the following perception scales are considered as continuous variables: one, the importance of acquisition cost in the choice of the main mode of urban transport (*aqc_cost*); two, the importance of operating cost (*op_cost*); three, the importance of safety in terms of accidentality (*accidents*); four, the importance of safety in terms of crime (*insecurity*);five, the importance of material convenience or comfort ($comfort$);six, the importance given to the level of air pollutant emissions; seven, the importance of travel time (*time*); eight, the perception of relatives (fam_perc); nine, the perception of close friends ($friend\_perc$); ten, the importance given to the perception of relatives ($impor\_fam$);and eleven, the importance given to the perception of friends ($impor\_friend$).

For each continuous variable considered, Table 4 provides the following information about its distribution: a skewness measure, the kurtosis measure and the p-value associated with the Shapiro-Wilk normality test. In general, if $skewness > 0 (skewness < 0)$, the distribution shows evidence of positive (negative) skewness. Likewise, if $kurtosis > 3 (kurtosis < 3)$, the



(a)



(b)

Figure 2: a) Communes in which users reside and b) communes to which users usually travel.

Table 4: Shapiro-Wilk normality test for continuous variables

	Kurtosis	*Skewness*	*W*	*p-value*
aqc_cost	2.54	-0.64	0.87	0^{***}
op_cost	2.48	-0.59	0.88	0^{***}
accidents	1.96	-0.25	0.89	0^{***}
insecurity	2.40	-0.73	0.83	0^{***}
comfort	2.50	-0.87	0.80	0^{***}
time	3.30	-1.18	0.76	0^{***}
pollution	2.12	-0.04	0.91	0^{***}
fam_perc	5.54	-1.76	0.64	0^{***}
friend_perc	5.17	-1.64	0.68	0^{***}
impor_fam	2.09	-0.60	0.83	0^{***}
impor_friend	1.53	0.18	0.84	0^{***}

Null hypothesis: the sample comes from a Gaussian distribution.
$^{***}p-value < 0.01$, $^{**}p-value < 0.05$, $^{*}p-value < 0.1$.

distribution is said to be leptokurtic (platykurtic) or, in other words, the distribution shows broad tails (the distribution shows high dispersion). In contrast, if $kurtosis = 3$, the distribution is meso-kurtic. Note that, according to Table 4, the continuous variables considered do not follow a normal distribution.

Considering the results of the Shapiro-Wilk normality test, Table 5 presents the median and the interquartile range associated with each. (As a complementary exercise, Table 12 presents an analogous descriptive summary reporting the mean and standard deviation). For the case of categorical variables, Table 2 presents the count and proportion of each category. Given a variable of k categorías, k categories, the Kruskal-Wallis nonparametric test is implemented to determine whether the k ppopulations are identical (see Table 7). Note that, based on the results of the Kruskal-Wallis non-parametric test for k categories, we propose the implementation of the Mann-Whitney non-parametric test according to pairs of categories. (For results differentiated according to pairs of socioeconomic strata, see Table **??**).

Finally, in a similar vein, in order to determine the independence between categorical variables, Table 8 presents the results of the implementation of the χ^2 test of independence; and Table 9, the results of the implementation of Fisher's exact test. In general terms, it is verified that the χ^2 test of independence and Fisher's exact test lead to identical conclusions.

3.4.2 Grouped variables

In the following, two grouped variables are considered: first, the time traveled daily, on average, by users of urban modes of transportation (*Time*); and second, the distance traveled daily, on average, from the participant's home to his or her frequent destination (*Distance*). The first variable considers six classes with amplitude of 9 minutes (i.e., 11 to 20 minutes, 21 to 30 minutes, 31 to 40 minutes, 41 to 50 minutes, 51 to 60 minutes and 61 to 70 minutes); the second, five classes with amplitude of 4 kilometers (i.e., 1 to 5 kilometers, 6 to 10 kilometers, 11 to 15 kilometers, 16 to 20 kilometers, 21 to 30 kilometers).

Table 5: Categorical variables and non-Gaussian continuous variables

	All	*Low*	*Medium*	*High*
accidents, Q2 (IQR)	3 (2, 4)	3 (3, 4)	3 (2, 4)	4 (2, 5)
comfort, Q2 (IQR)	4 (3, 5)	4 (2, 4)	4 (3, 5)	5 (4, 5)
aqc_cost, Q2 (IQR)	4 (3, 4)	4 (3, 5)	4 (3, 4)	4 (3, 4)
op_cost, Q2 (IQR)	4 (3, 4)	4 (3, 5)	4 (3, 4)	4 (3, 4)
insecurity, Q2 (IQR)	4 (3, 5)	3 (2, 5)	4 (3, 5)	5 (3, 5)
pollution, Q2 (IQR)	3 (2, 4)	3 (2, 4)	3 (2, 4)	3 (2, 4)
impor_friend, Q2 (IQR)	3 (1, 4)	3 (1, 4)	3 (1, 4)	3 (1, 5)
impor_fam, Q2 (IQR)	4 (3, 5)	3 (3, 5)	4 (3, 5)	4 (3, 5)
friend_perc, Q2 (IQR)	5 (4, 5)	4 (3, 5)	5 (4, 5)	5 (4, 5)
fam_perc, Q2 (IQR)	5 (4, 5)	5 (3, 5)	5 (4, 5)	5 (5, 5)
time, Q2 (IQR)	4 (3, 5)	4 (3, 5)	4 (3, 5)	5 (4, 5)
Gender, n (%)				
Female	277 (44.39%)	30 (42.25%)	149 (45.02%)	98 (44.14%)
Male	344 (55.13%)	41 (57.75%)	181 (54.68%)	122 (54.95%)
Prefer not to answer	3 (0.48%)	NA	1 (0.3%)	2 (0.9%)
Age, n (%)				
15 - 19 years	101 (16.19%)	14 (19.72%)	51 (15.41%)	36 (16.22%)
20- 24 years	179 (28.69%)	23 (32.39%)	90 (27.19%)	66 (29.73%)
25- 29 years	39 (6.25%)	8 (11.27%)	22 (6.65%)	9 (4.05%)
30- 34 years	57 (9.13%)	5 (7.04%)	37 (11.18%)	15 (6.76%)
35- 39 years	53 (8.49%)	2 (2.82%)	39 (11.78%)	12 (5.41%)
40 - 44 years	58 (9.29%)	7 (9.86%)	35 (10.57%)	16 (7.21%)
45- 49 years	47 (7.53%)	5 (7.04%)	16 (4.83%)	26 (11.71%)
50 - 54 years	34 (5.45%)	1 (1.41%)	16 (4.83%)	17 (7.66%)
55 - 59 years	35 (5.61%)	4 (5.63%)	15 (4.53%)	16 (7.21%)
60 -64 years	17 (2.72%)	1 (1.41%)	8 (2.42%)	8 (3.6%)
65 - 69 years	1 (0.16%)	1 (1.41%)	NA	NA
70 or older	3 (0.48%)	NA	2 (0.6%)	1 (0.45%)
Transport mode, n (%)				
App	42 (6.73%)	1 (1.41%)	23 (6.95%)	18 (8.11%)
Car	322 (51.6%)	10 (14.08%)	152 (45.92%)	160 (72.07%)
Motorcycle	100 (16.03%)	30 (42.25%)	60 (18.13%)	10 (4.5%)
Other	55 (8.81%)	3 (4.23%)	32 (9.67%)	20 (9.01%)
Taxi	1 (0.16%)	NA	NA	1 (0.45%)
Informal transport	11 (1.76%)	5 (7.04%)	5 (1.51%)	1 (0.45%)
Public transport	93 (14.9%)	22 (30.99%)	59 (17.82%)	12 (5.41%)

The category **Other** corresponds to bicycles, skateboards, walking, etc.; **Public transport**, , to formal public transport —MIO, e.g.—; and the category **App**, to applications such as Uber, Didi, Cabify, etc. The**Low** stratum includes strata 1 and 2; the **Medium** stratum, strata 3 and 4; and the **High** stratum, strata 5 and 6. *IQR* = Rango intercuartílico.

Table 6: Non-parametric Kruskal-Wallis test

	Gender		*Strata*		*Age*	
	$\chi^2(df)$	*p-value*	$\chi^2(df)$	*p-value*	$\chi^2(df)$	*p-value*
accidents	6.1 (2)	0.05^{*}	7.5 (2)	0.02^{**}	19.2 (11)	0.06^{*}
comfort	8 (2)	0.02^{**}	36.8 (2)	0.00^{***}	22.7 (11)	0.02^{**}
aqc_cost	0.3 (2)	0.84	0.8 (2)	0.66	16.4 (11)	0.13
op_cost	0.6 (2)	0.75	3 (2)	0.22	10.5 (11)	0.49
insecurity	14.1 (2)	0.00^{***}	31 (2)	0.00^{***}	11.4 (11)	0.41
pollution	7.4 (2)	0.02^{**}	1.6 (2)	0.46	47.3 (11)	0.00^{***}
impor_friend	4.4 (2)	0.11	3 (2)	0.23	11.8 (11)	0.38
impor_fam	1.7 (2)	0.44	5.3 (2)	0.07^{*}	21.1 (11)	0.03^{**}
friend_perc	3.1 (2)	0.21	27.6 (2)	0.00^{***}	8.1 (11)	0.70
fam_perc	0.2 (2)	0.92	43.3 (2)	0.00^{***}	15.6 (11)	0.16
time	2.7 (2)	0.26	9.3 (2)	0.01^{**}	39.1 (11)	0.00^{***}

Null hypothesis: The k-populations have identical distributions. df = degrees of freedom. $^{***}p-value < 0.01$, $^{**}p-value < 0.05$, $^{*}p-value < 0.1$.

Table 7: Non-parametric Mann-Whitney U test (Socioeconomic status)

	Low vs. Medium	*Low vs. High*	*Medium vs. High*
accidents	0.3197	0.4072	0.0065^{***}
comfort	0.0107^{**}	0.0000^{***}	0.0000^{***}
insecurity	0.1974	0.0000^{***}	0.0000^{***}
fam_perc	0.0649^{*}	0.0000^{***}	0.0000^{***}
friend_perc	0.0732^{*}	0.0000^{***}	0.0000^{***}

Null hypothesis: The two populations have identical distributions. $^{***}p-value < 0.01$, $^{**}p-value < 0.05$, $^{*}p-value < 0.1$.

Table 8: Chi-square (χ^2) test of independence

Z_1	Z_2	χ^2	*df*	*p-value*
gender	est	1.582724	4	0.8118937
gender	age	34.756655	22	0.0410731^{**}
gender	medium	10.109611	12	0.6063450
est	age	44.134795	22	0.0034115^{***}
est	medium	137.836147	12	0.0000000^{***}
age	medium	129.808251	66	0.0000047^{***}

Null hypothesis: the two variables are independent. $^{***}p-value < 0.01$, $^{**}p-value < 0.05$, $^{*}p-value < 0.1$.

Table 9: Fisher's exact test

Z_1	Z_2	*p-value*
gender	est	0.8628781
gender	age	0.0052386***
gender	medium	0.4432784
est	age	0.0061602
est	medium	0.0000001***
age	medium	0.0000001***

It is found that Chi-square (χ^2) test of independence and Fisher's exact test yield the same conclusion. $^{***}p - value < 0.01$, $^{**}p - value < 0.05$, $^{*}p - value < 0.1$.

Table 10: Variables by class intervals

	All	*Low*	*Medium*	*High*
Time, Q2 (IQR)	33.6 (21.6, 49.2)	47.5 (32.2, 62.4)	35.4 (23.6, 49.4)	26.8 (17.4, 41.2)
Time, n (%)				
11 - 20 minutes	147 (23.56%)	7 (9.86%)	62 (18.73%)	78 (35.14%)
21 - 30 minutes	131 (20.99%)	9 (12.68%)	71 (21.45%)	51 (22.97%)
31 - 40 minutes	116 (18.59%)	13 (18.31%)	66 (19.94%)	37 (16.67%)
41 - 50 minutes	81 (12.98%)	9 (12.68%)	53 (16.01%)	19 (8.56%)
51 - 60 minutes	69 (11.06%)	12 (16.9%)	39 (11.78%)	18 (8.11%)
61 - 70 minutes	80 (12.82%)	21 (29.58%)	40 (12.08%)	19 (8.56%)
Distance, Q2 (IQR)	9.5 (6.6, 16.4)	13.8 (9.4, 21.9)	9.7 (6.8, 16.8)	8.4 (4.5, 13.6)
Distance, n (%)				
1 - 5 kilometers	127 (20.35%)	4 (5.63%)	59 (17.82%)	64 (28.83%)
6 - 10 kilometers	210 (33.65%)	16 (22.54%)	115 (34.74%)	79 (35.59%)
11 - 15 kilometers	124 (19.87%)	22 (30.99%)	66 (19.94%)	36 (16.22%)
16 - 20 kilometers	71 (11.38%)	6 (8.45%)	42 (12.69%)	23 (10.36%)
21 - 25 kilometers	92 (14.74%)	23 (32.39%)	49 (14.8%)	20 (9.01%)

The variable **Time** corresponds to the time traveled daily, on average, by users of urban modes of transport. The **Distance** variable expresses the average distance traveled daily. The **Low** stratum includes strata 1 and 2; the **Medium** stratum, strata 3 and 4; and the **High** stratum, strata 5 and 6.IQR =Interquartile range.

For simplicity, for each grouped variable, the median and interquartile range are reported (see Table 10). Given $p \in (0,1) \subset \mathbb{R}^+$, let X_p the percentile of order p; and let n be the sample size. Define $k \in \mathbb{Z}^{++}$ as the first class to fulfill $F_k > np$ where $F_k$expresses the cumulative frequency for the class k. Therefore, X_p is given by the following formula:

$$X_p = L_k + w \left[\frac{np - F_{k-1}}{f_k} \right], \tag{5}$$

where L_k is the lower bound of the class k; f_k, the frequency of class k; and w, the class width. (For details of the calculation of the median and interquartile range according to classes, see Ramachandran & Tsokos, 2009, pp. 30 ff.). (As a complementary exercise, Table 13 presents the calculation of the mean and standard deviation for grouped variables.)

3.5 Perception scales compared by pairs of observations

Consider the following general hypothesis: given two perception scales $\vec{X} \in \mathbb{Z}^n$ and $\vec{Y} \in \mathbb{Z}^n$ for a sample size n,the objective is to assess whether the population values of $\vec{X}$ are, according to their trend, higher than the population values of $\vec{Y}$In the context of the urban mobility dynamics survey, consider the following example: let $\vec{X}$ express the importance assigned to safety in terms of crime; and $\vec{Y}$, the importance assigned to safety in terms of accident rate. Thus, the initial hypothesis aims to assess whether, in contrast, participants attach greater importance to safety in terms of delinquency. Since perception scales are variables susceptible to be analyzed according to pairs of observations, the implementation of the Wilcoxon signed-rank test is proposed (For formal details, see Wilcoxon, 1945, 1957).

Let $\vec{X} = (X_1, \ldots, X_n) \in \mathbb{Z}^n$ and $\vec{Y} = (Y_1, \ldots, Y_n) \in \mathbb{Z}^n$ be samples of size $n \in \mathbb{Z}^{++}$ for two continuous variables. Consider the following n ordered pairs of observations: $(x_1, y_1), \ldots, (x_n, y_n)$. Whatever $i \in \{1, \ldots, n\}$, define $|d_i|$ to be the absolute value of the difference for the i-th pair of observations such that $|d_i| = |y_i - x_i|$. Further, suppose that $m \in \mathbb{Z}^{++}$ is the number of pairs of observations satisfying that $|d_i| \neq 0$. All in all, consider a one-sided test whose hypothesis is given by.

$$\begin{aligned} H_0 &: d_k \geq 0 \\ H_1 &: d_k < 0 \end{aligned} \tag{6}$$

where d_k expresses the median of d_i for $i \in \{1, \ldots, n\}$. The decision rule is as follows: $T > w_{1-\alpha}$. (Note that T is the test statistic; and $w_{1-\alpha}$, the value of the Wilcoxon signed rank table with parameters m and $1 - \alpha \in (0,1) \subset \mathbb{R}^{++}$). The implementation of the exercise is presented differentiated according to socioeconomic strata. For low-stratum participants, the results are reported in Table 11(a);for middle-stratum participants, in Table 11(b);); and for high-stratum participants, in Table 11(c).

4 Annex

Table 11: Wilcoxon (1945, 1957) signed rank test. (Note that $^{***}p - value < 0.01$, $^{**}p - value < 0.05$, $^{*}p - value < 0.1$)

(a). Low socioeconomic stratum

	acc.	comf.	aqc.	op.	insec.	poll.	I.fri.	I.fam.	P.fri	P.fam	time
acc.	1	0.34	0.94	0.87	0.6	0***	0***	0.5	1	1	0.93
comf.	0.66	1	0.96	0.94	0.72	0.01**	0.01**	0.67	1	1	1
aqc.	0.07*	0.04**	1	0.3	0.1	0***	0***	0.13	0.97	0.99	0.5
op.	0.13	0.06*	0.72	1	0.15	0***	0***	0.24	0.99	0.99	0.77
insec.	0.4	0.29	0.9	0.86	1	0***	0***	0.49	1	1	0.93
poll.	1	0.99	1	1	1	1	0.25	0.98	1	1	1
I.fri.	1	0.99	1	1	1	0.75	1	1	1	1	1
I.fam.	0.5	0.33	0.87	0.76	0.52	0.02**	0***	1	1	1	0.84
P.fri	0***	0***	0.03**	0.01**	0***	0***	0***	0***	1	0.73	0.04**
P.fam	0***	0***	0.01**	0.01**	0***	0***	0***	0***	0.28	1	0.02**
time	0.07*	0***	0.5	0.23	0.07*	0***	0***	0.17	0.96	0.98	1

(b). Medium socioeconomic stratum

	acc.	comf.	aqc.	op.	insec.	poll.	I.fri.	I.fam.	P.fri	P.fam	time
acc.	1	1	1	1	1	0.02**	0***	1	1	1	1
comf.	0***	1	0.05*	0.04**	0.04**	0***	0***	0.14	1	1	1
aqc.	0***	0.95	1	0.64	0.75	0***	0***	0.56	1	1	1
op.	0***	0.96	0.37	1	0.71	0***	0***	0.55	1	1	1
insec.	0***	0.96	0.25	0.29	1	0***	0***	0.41	1	1	1
poll.	0.98	1	1	1	1	1	0***	1	1	1	1
I.fri.	1	1	1	1	1	1	1	1	1	1	1
I.fam.	0***	0.86	0.44	0.45	0.59	0***	0***	1	1	1	1
P.fri	0***	0***	0***	0***	0***	0***	0***	0***	1	0.83	0***
P.fam	0***	0***	0***	0***	0***	0***	0***	0***	0.18	1	0***
time	0***	0***	0***	0***	0***	0***	0***	0***	1	1	1

(c). High socioeconomic stratum

	acc.	comf.	aqc.	op.	insec.	poll.	I.fri.	I.fam.	P.fri	P.fam	time
acc.	1	1	0.69	0.18	1	0***	0***	0.94	1	1	1
comf.	0***	1	0***	0***	0.18	0***	0***	0***	1	1	0.53
aqc.	0.31	1	1	0.04**	1	0***	0***	0.88	1	1	1
op.	0.82	1	0.96	1	1	0***	0***	0.98	1	1	1
insec.	0***	0.83	0***	0***	1	0***	0***	0***	1	1	0.81
poll.	1	1	1	1	1	1	0.15	1	1	1	1
I.am.	1	1	1	1	1	0.85	1	1	1	1	1
I.fri.	0.06*	1	0.12	0.02**	1	0***	0***	1	1	1	1
P.am	0***	0***	0***	0***	0***	0***	0***	0***	1	1	0***
P.fri	0***	0***	0***	0***	0***	0***	0***	0***	0***	1	0***
time	0***	0.47	0***	0***	0.19	0***	0***	0***	1	1	1

Table 12: Categorical variables and non-Gaussian continuous variables

	All	*Low*	*Medium*	*High*
accidents, mean (sd)	3.29 (1.33)	3.32 (1.33)	3.16 (1.32)	3.47 (1.32)
comfort, mean (sd)	3.8 (1.35)	3.24 (1.42)	3.67 (1.37)	4.17 (1.21)
aqc_cost, mean (sd)	3.53 (1.22)	3.59 (1.35)	3.53 (1.23)	3.51 (1.18)
op_cost, mean (sd)	3.49 (1.22)	3.52 (1.36)	3.53 (1.23)	3.4 (1.17)
insecurity, mean (sd)	3.74 (1.31)	3.37 (1.34)	3.59 (1.31)	4.09 (1.21)
pollution, mean (sd)	3.01 (1.25)	2.89 (1.37)	2.99 (1.28)	3.09 (1.18)
impor_friend, mean (sd)	2.81 (1.57)	2.72 (1.49)	2.72 (1.52)	2.96 (1.66)
impor_fam, mean (sd)	3.57 (1.42)	3.35 (1.42)	3.54 (1.36)	3.67 (1.51)
friend_perc, mean (sd)	4.38 (0.97)	4 (1.22)	4.29 (1)	4.62 (0.76)
fam_perc, mean (sd)	4.45 (0.94)	4.08 (1.13)	4.34 (0.98)	4.73 (0.73)
time, mean (sd)	4.02 (1.25)	3.61 (1.49)	4 (1.23)	4.18 (1.17)
Gender, n (%)				
Female	277 (44.39%)	30 (42.25%)	149 (45.02%)	98 (44.14%)
Male	344 (55.13%)	41 (57.75%)	181 (54.68%)	122 (54.95%)
Prefer not to answer	3 (0.48%)	NA	1 (0.3%)	2 (0.9%)
Age, n (%)				
15 - 19 years	101 (16.19%)	14 (19.72%)	51 (15.41%)	36 (16.22%)
20- 24 years	179 (28.69%)	23 (32.39%)	90 (27.19%)	66 (29.73%)
25- 29 years	39 (6.25%)	8 (11.27%)	22 (6.65%)	9 (4.05%)
30- 34 years	57 (9.13%)	5 (7.04%)	37 (11.18%)	15 (6.76%)
35- 39 years	53 (8.49%)	2 (2.82%)	39 (11.78%)	12 (5.41%)
40 - 44 years	58 (9.29%)	7 (9.86%)	35 (10.57%)	16 (7.21%)
45- 49 years	47 (7.53%)	5 (7.04%)	16 (4.83%)	26 (11.71%)
50 - 54 years	34 (5.45%)	1 (1.41%)	16 (4.83%)	17 (7.66%)
55 - 59 years	35 (5.61%)	4 (5.63%)	15 (4.53%)	16 (7.21%)
60 -64 years	17 (2.72%)	1 (1.41%)	8 (2.42%)	8 (3.6%)
65 - 69 years	1 (0.16%)	1 (1.41%)	NA	NA
70 or older	3 (0.48%)	NA	2 (0.6%)	1 (0.45%)
Transport mode, n (%)				
App	42 (6.73%)	1 (1.41%)	23 (6.95%)	18 (8.11%)
Car	322 (51.6%)	10 (14.08%)	152 (45.92%)	160 (72.07%)
Motorcycle	100 (16.03%)	30 (42.25%)	60 (18.13%)	10 (4.5%)
Other	55 (8.81%)	3 (4.23%)	32 (9.67%)	20 (9.01%)
Taxi	1 (0.16%)	NA	NA	1 (0.45%)
Informal transport	11 (1.76%)	5 (7.04%)	5 (1.51%)	1 (0.45%)
Public transport	93 (14.9%)	22 (30.99%)	59 (17.82%)	12 (5.41%)

The category **Other** corresponds to bicycles, skateboards, walking, etc.; **Public transport**, , to formal public transport —MIO, e.g.—; and the category **App**, to applications such as Uber, Didi, Cabify, etc. The**Low** stratum includes strata 1 and 2; the **Medium** stratum, strata 3 and 4; and the **High** stratum, strata 5 and 6. *sd* = Standard deviation.

Table 13: Variables by class intervals

	All	*Low*	*Medium*	*High*
Time, Mean (sd)	36.04 (16.97)	45.78 (17.24)	37.19 (16.32)	31.22 (16.26)
Time, n (%)				
11 - 20 minutes	147 (23.56%)	7 (9.86%)	62 (18.73%)	78 (35.14%)
21 - 30 minutes	131 (20.99%)	9 (12.68%)	71 (21.45%)	51 (22.97%)
31 - 40 minutes	116 (18.59%)	13 (18.31%)	66 (19.94%)	37 (16.67%)
41 - 50 minutes	81 (12.98%)	9 (12.68%)	53 (16.01%)	19 (8.56%)
51 - 60 minutes	69 (11.06%)	12 (16.9%)	39 (11.78%)	18 (8.11%)
61 - 70 minutes	80 (12.82%)	21 (29.58%)	40 (12.08%)	19 (8.56%)
Distance, Mean (sd)	11.33 (6.6)	14.97 (6.52)	11.6 (6.53)	9.76 (6.24)
Distance, n (%)				
1 - 5 kilometers	127 (20.35%)	4 (5.63%)	59 (17.82%)	64 (28.83%)
6 - 10 kilometers	210 (33.65%)	16 (22.54%)	115 (34.74%)	79 (35.59%)
11 - 15 kilometers	124 (19.87%)	22 (30.99%)	66 (19.94%)	36 (16.22%)
16 - 20 kilometers	71 (11.38%)	6 (8.45%)	42 (12.69%)	23 (10.36%)
21 - 25 kilometers	92 (14.74%)	23 (32.39%)	49 (14.8%)	20 (9.01%)

The variable **Time** corresponds to the time traveled daily, on average, by users of urban modes of transport. The **Distance** variable expresses the average distance traveled daily. The **Low** stratum includes strata 1 and 2; the **Medium** stratum, strata 3 and 4; and the **High** stratum, strata 5 and 6. *sd* = Standard deviation.

References

Agarwal, A., Ziemke, D., & Nagel, K. (2020). Calibration of choice model parameters in a transport scenario with heterogeneous traffic conditions and income dependency. *Transportation Letters.*, *21*.

Ahmed, F., Catchpole, J., & Edirisinghe, T. (2020). Understanding young commuters' mode choice decision to use private car or public transport from an extended theory of planned behavior. *Journal of Transportation Research Board*, *2675*(3), 200 - 211.

Cochran, W. G. (1977). *Sampling techniques.* New York: John Wiley / Sons.

Faboya, O. T., Ryan, B., Figueredo, G., & Siebers, P. O. (2020). Using agent-based modelling for investigating modal shift: The case of university travel. *Computers Industrial Engineering*, *139*, 106077.

Fan, A., & Chen, X. (2020). Exploring the relationship between transport interventions, mode choice, and travel perception: An empirical study in beijing, china. *International Journal of Environmental Research and Public Health.*, *17*, 4258.

Fu, X. (2021). How habit moderates the commute mode decision process: integration of the theory of planned behavior and latent class choice model. *Transportation*, *48*, 2681 - 2707.

Gadepalli, R., Tiwari, G., & Bolia, N. (2018). Role of user's socio-economic and travel characteristics in mode choice between city bus and informal transit services: Lessons from household surveys in visakhapatnam, india. *Journal of Transport Geography*, *88*, 102307.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, *30*(3), 607 - 610.

Ramachandran, K. M., & Tsokos, C. P. (2009). *Mathematical statistics with applications.* Amsterdam: Elsevier.

Ryan, J. (2020). Examining the process of modal choice for everyday travel among older people. *International Journal of Environmental Research and Public Health*, *17*, 691.

S., Z., Vabuolytè, V., Burinskienè, M., & Antuchevicienè. (2020). Assessment of sustainable mobility by mcdm methods in the science and technology parks of vilnius, lithuania. *Sustainability*, *12*, 9947.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 180–183.

Wilcoxon, F. (1957). Some rapid approximate statistical procedures. *Annals of the New York Academy of Sciences*, *56*(6), 808–814.