Winter 2020 Geog 111B Assignment 3
Choice of Traveling Modes between Two Major Cities
– A Case Study of Trips between Montreal and Toronto, Canada
Jiamin Tan

Introduction

What travelling mode will people choose between two major cities? What variables can influence people's choices? Using the dataset "ModeCanada: Mode Choice for the Montreal-Toronto Corridor" from the "mlogit" package in R, a NULL discrete choice model, two basic discrete choice modes, and a mixed logit are fitted to discover the relationship between different variables and mode choices. The study finds that without any influence of measured variables in the dataset, people are most likely to travel between Montreal and Toronto by car; people always prefer a mode that is cheaper and takes less time both in and out the vehicle; people living in urban environments are more likely to use public transportation including trains and buses, and people with higher income prefer trips by air and cars. Finally, a basic model considering the income and urban factor fits the dataset the best examined by likelihood ratio tests, and the coefficients of willingness to pay are also calculated for each model.

Descriptive Statistics

MC

N: 11116

	cost	dist	freq	income	ivt	ovt
Mean	75.23	341.55	10.12	54.52	189.40	67.07
Std.Dev	51.16	151.34	11.72	17.51	117.92	43.02
Min	10.38	107.00	0.00	5.00	40.00	0.00
Median	58.05	326.00	5.00	55.00	161.00	77.00
Max	222.90	983.00	45.00	70.00	670.00	205.00

Generated by summarytools 0.9.4 (R version 3.5.1)

Table 1 – Descriptive Statistics

Table 1 shows the descriptive statistics of variables used for this study. There are 11116 samples from the data used for this study. We can see from the table that the average cost of is 75.23 Canadian dollars, the average distance travelled is 341.55 kilometers. The average frequency is 10.12 with a standard deviation 11.72, so the frequency of for each individual responded to the survey are quite different. The mean income of the respondents is 54520 Canadian dollars with a standard deviation of 17510, so the difference of the income of the respondents is not too large. The average in vehicle time is 189.40 minutes, and the average out vehicle time is 67.07 minutes. Lastly, for the categorical variable d_urban, there are 2948 respondents living in urban environment, while others live in non-urban environment.

NULL Model

```
Coefficients :
                 Estimate Std. Error z-value  Pr(>|z|)    
air:(intercept)  0.808287   0.055877  14.465 < 2.2e-16 ***
bus:(intercept) -3.835131   0.319625 -11.999 < 2.2e-16 ***
car:(intercept)  1.006680   0.054306  18.537 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 2 – Coefficients of the NULL Model

The equations of the systematic utility based on the NULL model:

$$V(\text{train}) = 0$$
$$V(\text{air}) = 0.808287$$
$$V(\text{bus}) = -3.835131$$
$$V(\text{car}) = 1.006680$$

The probability of using each mode based on the NULL model:

$$P(\text{train}) = \frac{e^{0}}{e^{0}+e^{0.808}+e^{-3.835}+e^{1.007}} = 0.167$$

$$P(\text{air}) = \frac{e^{0.808}}{e^{0}+e^{0.808}+e^{-3.835}+e^{1.007}} = 0.373$$

$$P(\text{bus}) = \frac{e^{-3.835}}{e^{0}+e^{0.808}+e^{-3.835}+e^{1.007}} = 0.004$$

$$P(\text{car}) = \frac{e^{1.007}}{e^{0}+e^{0.808}+e^{-3.835}+e^{1.007}} = 0.456$$

Firstly, we fit a NULL model in order to see without any measured variables, what will be people’s choices to travel between Montreal and Toronto. The NULL model that we derived above only takes into account of the alternative specific constant (ASC) of each mode with the reference mode “train”. The model considers no known variables, such as cost or waiting time, but it only considers how the unknown factors to the researchers influence a person’s choice. The ASCs for each mode comparing to the “train” are 0, 0.808, -3.835, 1.007 respectively. We can also calculate the probability for a person to use a mode comparing to the mode “train”, and the probabilities are 0.167, 0.373, 0.004, 0.456, respectively for train, air, bus, and car. From these probabilities, we can see that, without considering any measured variables in the dataset, a person is most willing to the travel by cars comparing to take a train, and a person is not likely to choose to travel by bus comparing to traveling by train. In addition, traveling by air is slightly more favorable than taking a train. The result obeys common sense that both trains and bus are public transportation which are not popular in North America due to uncertainties. The log likelihood of the NULL model is -2903.4.

Basic Model Using freq + cost + ivt + ovt

```
Coefficients :
                     Estimate  Std. Error  z-value  Pr(>|z|)
air:(intercept)  1.93433073  0.39756499   4.8654 1.142e-06 ***
bus:(intercept) -5.85322844  0.34759357 -16.8393 < 2.2e-16 ***
car:(intercept) -1.67816021  0.22434623  -7.4802 7.416e-14 ***
freq             0.09401755  0.00464254  20.2513 < 2.2e-16 ***
cost            -0.04554646  0.00386511 -11.7840 < 2.2e-16 ***
ivt             -0.00997360  0.00072653 -13.7277 < 2.2e-16 ***
ovt             -0.04263010  0.00280722 -15.1859 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 3(a) – Coefficients of the Basic Model with freq, cost, ivt, ovt

The equations of the systematic utility based on the basic model using freq + cost + ivt + ovt:

V(train) = 0 + 0.094*freq – 0.046*cost – 0.010*ivt – 0.043*ovt

V(air) = 1.934 + 0.094*freq – 0.046*cost – 0.010*ivt – 0.043*ovt

V(bus) = -5.853 + 0.094*freq – 0.046*cost – 0.010*ivt – 0.043*ovt

V(car) = -1.678+ 0.094*freq – 0.046*cost – 0.010*ivt – 0.043*ovt

The probability of using each mode based on the basic model above:

P(train) =

$$\frac{e^{0 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt}}{e^{0 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{1.934 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{-5.853 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{-1.678+ 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt}}$$

P(air) =

$$\frac{e^{1.934 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt}}{e^{0 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{1.934 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{-5.853 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{-1.678+ 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt}}$$

P(bus) =

$$\frac{e^{-5.853 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt}}{e^{0 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{1.934 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{-5.853 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{-1.678+ 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt}}$$

P(car) =

$$\frac{e^{-1.678 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt}}{e^{0 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{1.934 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{-5.853 + 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt} + e^{-1.67+ 0.094*freq - 0.046*cost - 0.010*ivt - 0.043*ovt}}$$

After fitting the NULL model, we now build the basic model above that takes into account of the frequency, monetary cost, in vehicle time, and out vehicle time. All the coefficients estimated by the model are statistically significant at a significance level of 0.05 shown in Table 3(a). We can see that the coefficient of the frequency is positive (0.094), that means people tend to choose modes with higher frequency. The coefficient of the cost is negative (-0.046), that means people are more like to choose cheaper modes. People also tend to choose modes with less in vehicle time because the coefficient for “ivt” is negative (-0.010). People dislike modes with long out vehicle time even more, because the coefficient of “ovt” is more negative (-0.043). In general, the model makes sense because people like to choose modes with high frequency, low cost, and little time. The log likelihood of this model is -1983.3 which is greater than the log likelihood of the NULL model (-2903.4), and the p-value derived by the Chisquared test is smaller than 0.05, so we can conclude that it is the two models are statistically different, and this basic model is at least better than the NULL model.

Willingness-to-pay respect to: cost

	Estimate	Std. Error	t-value	Pr(>|t|)	
air:(intercept)	-42.469400	5.579077	-7.6123	2.687e-14	***
bus:(intercept)	128.511167	10.920971	11.7674	< 2.2e-16	***
car:(intercept)	36.845021	6.202710	5.9401	2.848e-09	***
freq	-2.064212	0.165742	-12.4544	< 2.2e-16	***
ivt	0.218976	0.028298	7.7381	9.992e-15	***
ovt	0.935970	0.101050	9.2624	< 2.2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 3(b) – Coefficients of WTP of the Basic Model with freq, cost, ivt, ovt

The willingness to pay coefficients in Table 3(b) are derived by calculating the coefficient of each variable divided by the coefficient of the variable “cost” in Table 3(a). All the coefficients in this table are significant at the significance value of 0.05. Specifically, people will be willing to pay 2.064 Canadian dollars for a one-unit increase in frequency of a mode. People will be willing to pay 0.219 Canadian dollars to decrease one minute of in vehicle time, and people will be willing to pay 0.936 Canadian dollars to decrease one minute of out vehicle time.

Basic Model Using freq + cost + ivt + ovt and income and urban

Coefficients:

	Estimate	Std. Error	z-value	Pr(>|z|)	
air:(intercept)	-0.37044888	0.47360020	-0.7822	0.434099	
bus:(intercept)	-3.78630313	0.70620690	-5.3615	8.255e-08	***
car:(intercept)	-2.04868945	0.28751636	-7.1255	1.037e-12	***
freq	0.08858758	0.00525928	16.8441	< 2.2e-16	***
cost	-0.04234811	0.00399465	-10.6012	< 2.2e-16	***
ivt	-0.01016126	0.00075999	-13.3702	< 2.2e-16	***
ovt	-0.04162458	0.00284927	-14.6088	< 2.2e-16	***
air:income	0.03672688	0.00400731	9.1650	< 2.2e-16	***
bus:income	-0.05025325	0.01791559	-2.8050	0.005032	**
car:income	0.00970083	0.00315540	3.0744	0.002110	**
air:d_urban	-0.01075577	0.15594352	-0.0690	0.945012	
bus:d_urban	0.41612559	0.70352677	0.5915	0.554195	
car:d_urban	-0.43377543	0.14233366	-3.0476	0.002307	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 4(a) – Coefficients of the Basic Model with freq, cost, ivt, ovt, income, d_urban

The equations of the systematic utility based on the basic model using freq + cost + ivt + ovt + income + d_urban:

V(train) = 0 + 0.089*freq – 0.042*cost – 0.010*ivt – 0.042*ovt

V(air) = -0.370 + 0.089*freq – 0.042*cost – 0.010*ivt – 0.042*ovt + 0.037*income – 0.011 *d_urban

V(bus) = -3.786 + 0.089*freq – 0.042*cost – 0.010*ivt – 0.042*ovt – 0.050*income + 0.416 *d_urban

V(air) = -2.049 + 0.089*freq – 0.042*cost – 0.010*ivt – 0.042*ovt + 0.010*income – 0.434 *d_urban

The probability of using each mode based on the basic model above (the full equations are too long to list, so utility functions from above will be used directly in the equations below):

$$P(\text{train}) = \frac{e^{v(\text{train})}}{e^{v(\text{train})}+e^{v(\text{air})}+e^{v(\text{bus})}+e^{v(\text{car})}}$$

$$P(\text{air}) = \frac{e^{v(\text{air})}}{e^{v(\text{train})}+e^{v(\text{air})}+e^{v(\text{bus})}+e^{v(\text{car})}}$$

$$P(\text{bus}) = \frac{e^{v(\text{bus})}}{e^{v(\text{train})}+e^{v(\text{air})}+e^{v(\text{bus})}+e^{v(\text{car})}}$$

$$P(\text{car}) = \frac{e^{v(\text{car})}}{e^{v(\text{train})}+e^{v(\text{air})}+e^{v(\text{bus})}+e^{v(\text{car})}}$$

We now add more variables to our basic model. Table 4(a) shows the coefficients of the basic model take into account the frequency, monetary cost, in vehicle time, out vehicle time, income, and d_urban. Specifically, higher frequency will increase the utility, because frequency has a positive coefficient (0.089). According to this model, people also prefer to modes with lower cost, shorter time both in and out of the vehicles, the coefficients are -0.042, 0.012, -0.042 respectively. In terms of the influence of income on utility, we can see that the coefficient of choosing bus is negative (-0.050). That means people with higher income prefer not to use buses. We can also see that people with more income are more likely to travel by air rather than car (0.037 vs 0.010). Train is also not a preferable option for people high income (0). This obeys common sense that people with higher income don't like public transportation and they can afford for the air trips. In terms of whether a person living in urban or non-urban environment will affect the utility, we can see that people living in urban area prefer bus rather than the other three modes, because the "bus" variable has a positive coefficient (0.416), while other variables has negative or zero coefficients, -0.011, -0.434, and 0 respectively to air, car, and train. We also can conclude that people who live in non-urban environment prefer to use cars, since the coefficient is the most negative. The log likelihood of this model is -1923.6 which is greater than the log likelihood of the previous model (-1983.3). The p-value derived from the Chi-square test is smaller than 0.05, so we can conclude that this model is significantly different from the model above, and it is better than the previous model.

Willigness-to-pay respect to: cost					
	Estimate	Std. Error	t-value	Pr(>|t|)	
air:(intercept)	8.747708	11.858434	0.7377	0.460710	
bus:(intercept)	89.409030	17.162969	5.2094	1.894e-07	***
car:(intercept)	48.377357	8.418746	5.7464	9.117e-09	***
freq	-2.091890	0.191080	-10.9477	< 2.2e-16	***
ivt	0.239946	0.033629	7.1351	9.670e-13	***
ovt	0.982915	0.114037	8.6193	< 2.2e-16	***
air:income	-0.867262	0.126083	-6.8785	6.048e-12	***
bus:income	1.186671	0.438774	2.7045	0.006840	**
car:income	-0.229073	0.077034	-2.9737	0.002943	**
air:d_urban	0.253985	3.684627	0.0689	0.945045	
bus:d_urban	-9.826309	16.624272	-0.5911	0.554465	
car:d_urban	10.243089	3.661171	2.7978	0.005146	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4(b) – Coefficients of WTP of the Basic Model with freq, cost, ivt, ovt, income, d_urban

Table 4(b) shows the willingness to pay coefficients derived by calculating the coefficient of each variable divided by the coefficient of monetary cost. We can see that this model estimates that people will be willing to pay 2.092 Canadian dollars for one-unit increase in frequency; people will be willing to pay 0.240 Canadian dollars for one minute decrease in in vehicle time, and 0.983 Canadian dollars for one minute decrease in out vehicle time. Therefore, it indicates that people feel more satisfied when they are in a vehicle, that makes sense because nobody like to wait outside a vehicle.

Mixed Logit Model with Random In Vehicle Time

Coefficients:					
	Estimate	Std. Error	z-value	Pr(>|z|)	
air:(intercept)	2.2614835	0.5032909	4.4934	7.010e-06	***
bus:(intercept)	-7.3295901	0.4601962	-15.9271	< 2.2e-16	***
car:(intercept)	-2.0659557	0.2769656	-7.4593	8.704e-14	***
freq	0.1583464	0.0139716	11.3335	< 2.2e-16	***
cost	-0.0674646	0.0057737	-11.6848	< 2.2e-16	***
ovt	-0.0532017	0.0036928	-14.4070	< 2.2e-16	***
ivt	-0.0140146	0.0014265	-9.8242	< 2.2e-16	***
sd.ivt	0.0126561	0.0016197	7.8136	5.551e-15	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 5(a) – Coefficients of the Mixed Model

The equations of the systematic utility based on the mixed model using random ivt:

$$V(train) = 0 + 0.158*freq – 0.067*cost – 0.053*ovt + \beta*ivt$$

$$V(air) = 2.261 + 0.158*freq – 0.067*cost – 0.053*ovt + \beta*ivt$$

$$V(bus) = -7.330 + 0.158*freq - 0.067*cost - 0.053*ovt + \beta*ivt$$
$$V(car) = -2.066 + 0.158*freq - 0.067*cost - 0.053*ovt + \beta*ivt$$

The probability of using each mode based on the mixed model:

$$P(train) = \int \left(\frac{e^{v(train)}}{e^{v(train)}+e^{v(air)}+e^{v(bus)}+e^{v(car)}}\right) f(\beta)\, d\beta$$
$$P(air) = \int \left(\frac{e^{v(air)}}{e^{v(train)}+e^{v(air)}+e^{v(bus)}+e^{v(car)}}\right) f(\beta)\, d\beta$$
$$P(bus) = \int \left(\frac{e^{v(bus)}}{e^{v(train)}+e^{v(air)}+e^{v(bus)}+e^{v(car)}}\right) f(\beta)\, d\beta$$
$$P(car) = \int \left(\frac{e^{v(car)}}{e^{v(train)}+e^{v(air)}+e^{v(bus)}+e^{v(car)}}\right) f(\beta)\, d\beta$$

Lastly, we fit a mixed logit model to see if this model is better than other models above. From Table 5(a) above, all the coefficients are significant at the significance level of 0.05. We know that since the coefficient of the frequency variable is positive (0.158), a mode utility will increase if the it has a higher frequency. We can also see that, similar to all the basic models above, the coefficients of monetary cost and out vehicle time are negative (-0.067, - 0.053). Therefore, people will prefer to choose modes that have lower cost and shorter time out of the vehicles. The coefficient of the random "ivt" variable tells us that since it is negative, on average, people are less likely to choose a mode with high in vehicle time. Since the standard deviation (0.013) of the random "ivt" is close to the absolute value of the mean value (0.014), we can conclude that respondents in this dataset are very different from each other in terms of travelling mode choices. The log likelihood of this mixed model is -1948.8 which is smaller than the log likelihood of the previous model (-1923.6). The p-value derived from the Chi-square test is smaller than 0.05, so we can conclude that the mixed model is significantly different from the previous basic model taking into account of income and urban factor, and the previous model is better than this mixed model since it has a greater log likelihood value. Based on all the discussion of the log likelihood of all the model fitted, we conclude that the basic model with income and urban factor is the best model for calculating the utility value.

```
Willigness-to-pay respect to:  cost

                  Estimate Std. Error  t-value  Pr(>|t|)    
air:(intercept) -33.521020   5.600529  -5.9853 2.160e-09 ***
bus:(intercept) 108.643436   7.343698  14.7941 < 2.2e-16 ***
car:(intercept)  30.622794   4.959478   6.1746 6.633e-10 ***
freq             -2.347102   0.173925 -13.4949 < 2.2e-16 ***
ovt               0.788586   0.077191  10.2160 < 2.2e-16 ***
ivt               0.207732   0.027886   7.4493 9.370e-14 ***
sd.ivt           -0.187596   0.035351  -5.3067 1.116e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 5(b) – Coefficients of WTP of the Mixed Model

Table 5(b) shows the coefficients of willing to pay derived by the mixed model. It gets similar results comparing to the basic models. People will be willing to pay 2.347 Canadian dollars if there will be a one-unit increase in mode frequency; people will be willing to pay 0.789 Canadian dollars if the out vehicle time decreases by one minute; and people will be willing to pay 0.208 Canadian dollars if the in vehicle time decreases by one minute.

Conclusion

Based on the likelihood ratio tests conducted, we can conclude that the basic model taking into account the income and urban factors is best model. We can also conclude from the model that people tend to choose the traveling modes with higher frequency, less cost, and less time both in and out of the vehicle. In addition, the we can find that people with higher income tend to use air and car trips, and people who live in urban areas tend to use public transportation to travel between Montreal and Toronto.

Appendix

tan_j_lab3

Jiamin Tan

March 8, 2020

**Codes used in this lab are based on GauchoSpace handout

```
library(mlogit)
library(stargazer)
library(gmnl)
library(tidyverse)
library(summarytools)

# data
data("ModeCanada", package = "mlogit")
MC <- mlogit.data(ModeCanada, subset = noalt == 4, chid.var = "case", alt.var
= "alt", drop.index = TRUE)

# dummy variable for urban
MC$d_urban <- ifelse(MC$urban =='2', 1, 0)
MC$d_urban <- ifelse(is.na(MC$d_urban), 0, MC$d_urban)

# descriptive statistics
stats_table <- descr(MC[,c(2:7)], stats = c("mean", "sd", "min",
                                            "med", "max"))
view(stats_table)
sum(MC[,10])

# fit the null model
Null.CA.MNL <- mlogit(choice ~ 0 | 1, data=MC)
summary(Null.CA.MNL)

# fit the basic model w/o urban and income
Basic.CA.MNL <- mlogit(choice ~ freq + cost + ivt + ovt | 1, data=MC)
summary(Basic.CA.MNL)

# Use `gmnl` to get the willingness to pay
Canada.basic <- gmnl(choice ~ freq + cost + ivt + ovt | 1, data = MC, model="
mnl")
wtp.gmnl(Canada.basic, wrt = "cost")
# likelihood ratio test
lrtest(Null.CA.MNL,Canada.basic)

# fit the basic model w/ urban and income
Canada.basicplus <- gmnl(choice ~ freq + cost + ivt + ovt | 1 + income + d_ur
ban,data = MC, model="mnl")
summary(Canada.basicplus)
```

```r
wtp.gmnl(Canada.basicplus, wrt = "cost")
# likelihood ratio test
lrtest(Canada.basicplus, Canada.basic)

# fit the mixed model
Canada.mixl <- gmnl(choice ~ freq + cost + ivt + ovt | 1 ,
                    data = MC, model = "mixl", ranp = c(ivt = "n"), R = 200)
summary(Canada.mixl)
wtp.gmnl(Canada.mixl, wrt = "cost")

# likelihood ratio test
lrtest(Canada.basicplus, Canada.mixl)
```