Investigating students preferences on *Pasto Lesto* with choice-based conjoint analysis

Laboratory of Customer and Business Analytics

Letizia Monti
University of Trento
student ID: 203960
letizia.monti@studenti.unitn.it

Bruno Papa
University of Trento
student ID: 205832
bruno.papa@studenti.unitn.it

Abstract

The present research investigates students preferences on Pasto Lesto, a fixed menu meal offered by the food service of the University of Trento using choice based conjoint-based analysis. The survey design is fully randomized and the data were collected using a chatbot that asked respondents questions on Pasto Lesto preferences and on individual characteristics in a conversational manner that provided an engaging user experience. Data have been analysed using multinomial logistic model and mixed multinomial mixed model to capture the heterogeneity of the preferences of the different students.

1. Introduction

Understanding customers' behaviours and preferences has become an important goal in order to assess customers' satisfaction and improve businesses tailoring their products and services to the customers' needs and preferences. This is what have always been done by marketing divisions in order to increase the number of customers and the satisfaction of the existing ones towards the offered products. From another point of view marketing strategies are useful because they allow the creation of products more and more based on the actual needs and taste of the target population. Investigating customers' preferences require understanding how individuals process the available information in order to form an opinion toward a product and make a choice among various alternatives. In this sense, individuals' choices are considered as a proxy of their preferences.

The present research aims at analysing students' preferences towards *Pasto Lesto*, a particular fixed menu meal provided by the food service in the canteens of the University of Trento. The *Pasto Lesto* option provides reduced portions of a first course, a second course and a side dish from a fixed menu (i.e. fixed quantity and restricted possibility to choose among different dishes), and one element among fruit, yogurt or dessert. *Pasto Lesto* comes at the fixed price of €3,10. It was created few years ago with the aim of speeding up the service. Students have often few time to eat and *Pasto Lesto* offers significantly lower waiting times, thus it is an important option for the student's meal. Many times, the quantities of the elements in the served dish do not really seem satisfactory, a problem that hinders the reputation of this otherwise convenient option.

The aim of our investigation is to understand students' preferences towards the components of the ”Pasto Lesto” meal in order to discover which are the most important attributes and how the preference of these attributes vary across respondents. This in turn can be useful to improve the service offered by matching both the taste of consumers and the food service needs of providing a complete meal with at least a first course, a second course and a side dish at a fixed low price.

In particular, in the following sections it will be described some methodological aspects related to the techniques used, the survey design adopted to collect the data, the models used for the analysis and the obtained results.

2. Methodology

In order to investigate the preferences of the students with respect to the composition of the *Pasto Lesto*, we decided to use the framework of choice-based conjoint analysis. This technique emulates consumers' choices by presenting to the respondents a set of alternatives made of attributes with different levels among which to choose. The method is based on the assumption that an individual makes a choice from a set of alternatives so that to maximise his utility. With respect to the traditional conjoint analysis it allows to discover which are the relevant attributes used to make a choice and how these vary across respondents according to other individual specific characteristics.

In addition, choice-based conjoint analysis can be used to assess the customers' sensitivity to changes in the level of the attributes, to interpret the ratio of the coefficient of attributes and price to estimate respondents willingness to pay (*wtp*), and to compare a limited number of attribute profiles by simulating preferences shares in order to discover which will be the preferred profiles according to the investigated population.

We used this framework mainly because the task of choosing is natural and seemed to be more engaging for the respondents. Moreover, the students are required to choose at least part of the meal as in some occasion they can also decide to give up one of the meal component in order to have more of another dish. However, this last option is not really standardized and depends on individual practices of the canteen, the menu of the day and the particular canteen worker.

For the survey we did not provide any monetary incentive, instead we relied on other motivation factors such as "Altruism" which can be strong inside academic communities (as the one targeted by our investigation). In addition, the topic studied constitutes a relevant part in the life of many students, thus we tried to explain the aim of our experiment with the hope that students saw it as a possible mean of improving that aspect of their life, as much as we do.

Finally, to deliver the survey we decided to use a chatbot on the telegram platform, an instant messaging service. At present day, *Telegram* is one of the most famous messaging services, and although it has not the same diffusion of *Whatsapp*, it provides several advantages which have boosted its spreading among young people. For this reason we thought that the audience reached through this tool would be nearly as large as the one obtained through web sites, at least in this particular case, in which we target young students.

This choice allowed us to personalize our surveys: the appearance but also the techniques with which we delivered the questions allowing us to use open source alternatives to the paid tools.

We aimed also at providing a more engaging user experience: the user would not be pressed to complete the survey because it has opened an ad-hoc web page on the browser and can't do anything else with his smartphone. Instead, our survey will be just one of his chats and he could complete the survey whenever he wants, in a conversational manner.

The investigation has been spread through word of mouth in university, *facebook* groups of students and by means of flyers in the bulletin boards of the university departments and canteens. The service has been maintained online for two weeks: from the 9th to the 20th of December 2019.

3. Survey design

The application of choice-based conjoint analysis required the collection of data based on the choices of profiles among alternatives from a set choice, by respondents. In order to do that, the survey was designed in three stages: the selection of the attributes and levels for the *Pasto Lesto* profiles; the generation of the questions provided to each respondent; and the creation of the tool through which deliver the survey to the target population.

Starting from all the components of the *Pasto Lesto* provided by the food service, seven relevant attributes were identified as follows:

- First course: 60g, 80g, 100g
- Second course: 60g, 80g, 100g
- Side dish: 60g, 80g, 100g
- Fruit: yes, no
- Dessert: yes, no
- Coffee: yes, no
- Price: €2.50, €2.80, €3.10, €3.40, €3.70

where the respondents were told to regard 80g as the standard portion for the first and second course and for the side dish.

The levels of the attributes were chosen in order to represent variations in the quantities of food provided for each course (i.e. a smaller and larger portion than the standard one), the presence of extras included (i.e. fruit, dessert, coffee) and variations in the price (i.e. two lower and two higher prices than the current one).

In addition, three questions about characteristics of the individuals were devised about sex, frequency with which they eat at the university canteens and frequency with which they go back home (interpreted as stable place of residence):

- Sex: female, male, other
- Canteen frequency: daily, weekly, monthly, yearly
- Home return frequency: daily, weekly, monthly, yearly.

3.1. Alternatives generation

A fully randomized design was used for the survey in order to give to each respondent a random set of alternative profiles. The total choice set, made of ($3^3 \cdot 2^3 \cdot 5 = 1080$) possible profiles, was created by including all the possible combinations of the levels of each attribute. Respondents were asked 7 questions in which they had to choose the preferred *Pasto Lesto* profile among 3 alternatives.

Each question related to the *Pasto Lesto* choice was created by independently sampling with replacement three profiles from the entire choice set, in order to give to every respondent a random set of alternatives made of a random combination of attributes. In this way, repetitions of the same profile within the same questionnaire would have been possible, even though with a low probability ($P \leq \frac{1}{540}$), because of the large number of profiles in the choice set.

3.2. Survey delivery

The survey delivery was performed through a chatbot implemented on the *Telegram* messaging platform as previously described in the methodology section.

The flow of the user experience was created as follows (every step can be paused and resumed after closing the application as well):

- The user obtains the address/ID of the chatbot.
- The user has the ID of the chatbot, and opens the relative chat.
- A short description is visible on what the chatbot does, who is performing the study and what is the aim of the project.
- The user has to click the *start* button if he wants more information about what will be seen in the survey.
- At the end of the additional description, to start the survey the user has to push a button to declare that it allows us to process his data for the finalities of the described project.
- The bot sends a message with the first set of alternative options for the meal; the user can see only one choice set at the time and old and future questions are not visible. The optional element(coffee, dessert and yogurt) are displayed only if present in the configuration (i.e. displayed if value = yes, otherwise they are not printed in order to simulate absence in a real setting and to reduce complexity of the configurations when possible).
- A notification lets the user know that he is middle way from completing the survey.
- Once the questions about the meal are completed, the bot asks for the last three questions on additional variables.
- When all questions are answered a greeting message is displayed, closing the interaction.

3.3. Dataset

The total number of respondents after the investigation period was 142. Among them, 127 have completed the survey and 15 did not answer to one or more question. From the survey design, the questions about individual characteristics are the last three ones to be presented so, for the incomplete surveys, these information would be missing. Because the incomplete surveys are a small fraction and because we want to study also the relationship between preferences and individual characteristics, we chose to discard data of incomplete surveys from the analysis.

Data were organised according to the *Long* format required for the analysis, with all the alternatives for each question related to the "Pasto Lesto" in a row. The resulting dataset was made of 2655 rows and 14 variables:

1. User ID
2. Question number $\{1 \div 7\}$
3. Alternative number $\{1 \div 3\}$
4. Choice outcome $\{1, 0\}$
5. Attributes of the questions posed: first course, second course, side dish, fruit, dessert, coffee, price
6. Individual variables: sex, home visitation frequency, canteen visitation frequency.

4. Preliminary and exploratory data analyis

In this section we will have a preliminary look at the dataset, in particular at the individual specific variables we collected: sex, back home frequency and canteen meal frequency.

We already clarified that we have a total of 127 respondents who completed our survey and the following data will be drawn from that reduced dataset. Firstly, from Figure 1 we observe that our sample is nearly equally represented in terms of gender (the precise numbers are in the caption of the figure).

Then, from Figure 2, we can observe that the majority of the respondents return back home on a weekly basis, so probably are students from neighbouring regions who stay in Trento to avoid a possible long daily commute. At the second place are the ones who return home on a daily basis, probably resident in Trento or in a place not too far to reach in terms of commuting time. Finally, the tail is constituted by the ones who return home only on a monthly or yearly basis.

From the bar plot in Figure 3 referred to the canteen visitation frequency we can observe that the distribution is similar to the one in Figure 2. Weekly frequency is the most voted, but this time it is followed by monthly frequency and, daily and yearly frequency are at the same level.

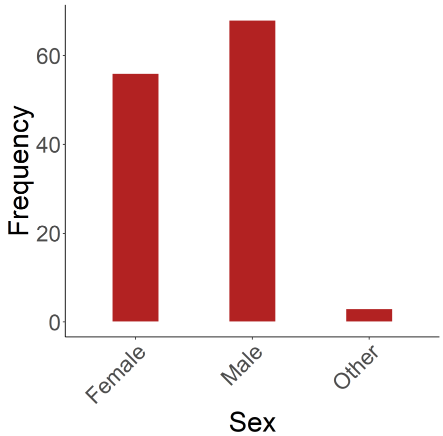

Figure 1. F:58, M:68, O:3.

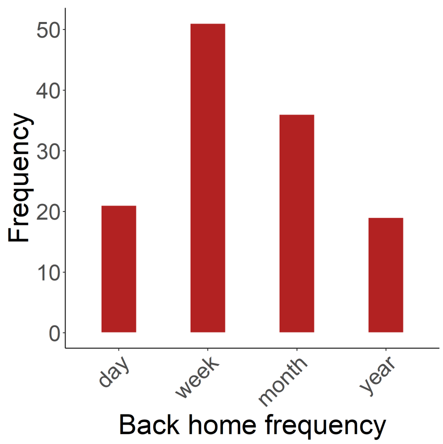

Figure 2. D:21 W:36 M:51 Y:19.

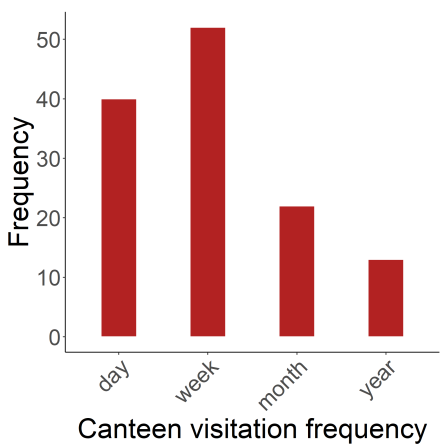

Figure 3. D:52 W:40 M:22 Y:13.

Now for a final data exploration we want to observe how the attitudes towards going back home and going to the canteens are related.

From the heatmap in Figure 4 we can observe that, as ex-

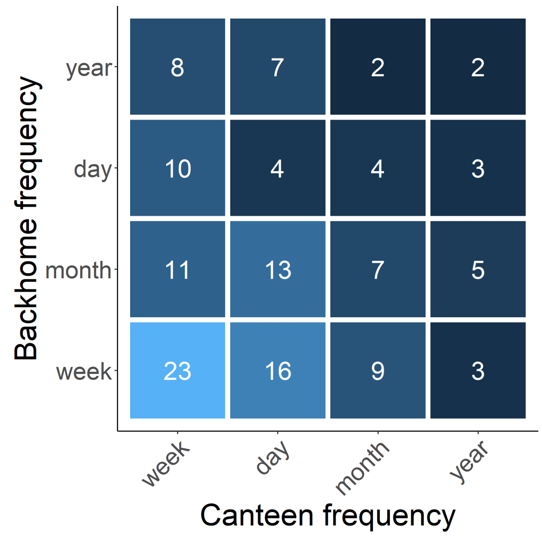

Figure 4. Heatmap for the relation between backhome frequency and canteen frequency.

pected, the most frequent combination is the one of the people who use the canteen at least one time per week and also go back home on a weekly basis. Then, the second most frequent combination regards those that use the canteen daily and return back home on a weekly basis. In general, we can observe that the majority of the respondents that use the canteens at least once per day or week (high frequency) are also the ones that go home once per week or month (mid frequency).

5. Choice-based conjoint analysis

From our dataset we decided initially to use all variables as categorical variables with the exception of price which preserved its numeric value. It is implicit that in the following analysis the computations were performed using *R v3.6.1* and the package *mlogit v1.0-1* for fitting multinomial logistic models.

5.1. Multinomial logistic model

We started to investigate part worths arising from all the dataset, without considering the nested structure deriving from different users. In order to do that, we firstly had to convert our data into a suitable format to work with the *mlogit* library, so we used the *mlogit.data* function from the *mlogit* package, then we were ready to start.

A first, simple multinomial logistic model, fitted with all the variables gives us the coefficients in Table 1. Here the estimates tell us how much each attribute is associated with the choices of the respondents. In particular, we can see that the intercepts are not significant which is a good sign and inform us that no particular position or buttons were preferred solely by its physical position.

	Estimate	Error	p-value
2:(intercept)	3.9982e-02	9.3550e-02	0.6690940
3:(intercept)	7.5533e-05	9.4529e-02	0.9993625
first.course60g	-1.0503e+00	1.2008e-01	**< 2.2e-16**
first.course80g	-2.0193e-01	1.0696e-01	**0.0590390**
second.course60g	-7.8836e-01	1.1500e-01	**7.121e-12**
second.course80g	-3.5858e-01	1.1508e-01	**0.0018332**
side.dish60g	-4.4863e-01	1.1678e-01	**0.0001222**
side.dish80g	-1.3583e-01	1.1427e-01	0.2345647
fruityes	4.9870e-01	9.3438e-02	**9.437e-08**
dessertyes	4.4742e-01	9.4763e-02	**2.341e-06**
coffeeyes	2.5350e-02	9.4390e-02	0.7882636
price	-1.6616e+00	1.2402e-01	**< 2.2e-16**

Table 1. Coefficients of the MNL model.

For what concerns the attributes, we can notice that there are some not really statistically significant. For example, the level 80g for both the first course and the side dish seem not to be a significant alternative with respect to the base level, in this case 100g.

The non significance of the level *coffe* $= yes$ could make us think that coffee does not play a significant role in the decision process of the respondents. Unfortunately, it is too early to confirm this conclusion, as we have not considered that we are dealing with different respondents yet and that they may have different attitudes toward the presence of coffee in the *Pasto Lesto*.

One last observation about this first model is that for some parameters the standard errors are quite high. The reasons for this can be many: the number of respondents not too large or the way in which we delivered questions. Maybe encouraging the respondent to answer to more different configurations could be an improvement. In any case, the signs of the coefficients are mostly consistent with our expectations, so we are confident in the fact that fixing the two issues above would produce more accurate results that would still go in the same direction.

Refining the model We want to check whether we can drop some irrelevant parameters, first of all the intercepts. We compare the two nested models, with and without the intercepts, using the Likelihood Ratio test. It can be done in R using the function *lrtest()* from *mlogit* package. This test confirmed that it is possible to drop the intercepts without significant loss of information.

Now our model is quite minimalistic, we are still not considering any interaction, thus we want to try a different method to interpret the model parameters. As we are considering price as a quantitative variable, a viable option is to consider the ratio of the coefficients on the price coefficient and interpret this object as the average respondent *willingness to pay* (wtp):

$$wtp_i = \beta_i/\beta_{price} \tag{1}$$

Parameter	WTP
first.course60g	0.63
first.course80g	0.12
second.course60g	0.47
second.course80g	0.22
side.dish60g	0.27
side.dish80g	0.08
fruityes	-0.30
dessertyes	-0.27
coffeeyes	-0.01

Table 2. Willingness to pay for the MNL model.

Table 2 shows the *wtp* from this model, from which we can get a better interpretation of our data. We can state for example that the average respondent is willing to pay at most €0.63 to pass from a configuration of the meal with a 60g first course to one with a 100g first course. Or equivalently, €0.63 would be the price at which the choice between 100g or 60g of first course become irrelevant.

Then, as we had only few levels for the meals and the price variable, but they are inherently quantitative, we want to check whether we can continue to use the meals as categorical variables and the price as quantitative. In order to do that we firstly compare the goodness of fit of the present model with respect to the same model but considering price as a categorical variable with 5 levels. The Likelihood Ratio test shows that using price as a factor significantly improves the model, so we will use it as a factor variables from now on.

Finally we compared our model to one in which the first course, the second course and the side dish were numeric variables, with their values in grams. Performing a test on this two model shows that considering meals as quantitative lead to a slightly worst goodness of fit, a difference not significant at the 95% confidence level. Here we could choose to accept the worst quality of the fit in order to reduce the degrees of freedom of our model; on the other hand the information that 80g of the first course or side dish is less significant in the choice, with respect to their 60g level, suggested that the relation between the quantity of food and choice probability is not uniform, so we decided to maintain food quantities as factor variables. It could be a reasonable hypothesis that the asymmetry in the coefficients for the various levels of the food quantities and price is given by their low value on this scale, identifiable as inherently different quantities and not as values on a common scale where the difference is proportional to their numeric values.

In conclusion, we will perform the rest of the analysis using only factor variables.

Predicting preference shares Now that we have our model fitted on the dataset, one interesting use is to predict the *preference shares* of a selected pool of alternative configurations of the *Pasto Lesto*. In practice, given the choice made by our population and a set of k different profiles, we can use our model to simulate the shares with which each of the k configuration would be preferred if this choice was to be presented to an audience.

Now we wanted to design a meal configuration from our experience and then use this tool in order to see how well it would go with respect to other possible configurations. Suppose that we want to give up a little quantity of the first course, gain a little of the second course, with same quantity of side dish as the standard *Pasto Lesto*. For the extras, we choose to offer only fruit, favouring the food service by avoiding to offer also yogurt, dessert or coffee and reducing the costs. Predicting this theoretical choice with respect to other possible configurations using the model fitted on our respondents we get the results of Table 3.

Share	First C.	Second C.	Side d.	fruit	dessert	coffe	price
19 %	60	80	100	no	no	no	€2.5
9 %	80	80	100	no	yes	yes	€3.7
33%	**60**	**100**	**80**	**yes**	**no**	**no**	**€2.8**
14 %	100	100	80	no	yes	no	€3.7
22 %	60	100	80	no	yes	no	€3.1

Table 3. Preference shares prediction with MNL model.

It shows that our configuration would be quite successful having the largest preference share. In addition, we can notice that it is more preferred than the configuration with the lower price, confirming that price *per se* is not the most influential factor.

5.2. Mixed multinomial logistic model

The mixed multinomial logistic model allows the inclusion of both individual and alternative specific variables, by considering the heterogeneity of the population on hand.

In practice, it is assumed that each respondent has his own coefficients with their relative standard deviations from the average coefficients values of the entire sample. Assuming that the distribution of the random parameters is Normal, the model is fitted with all the parameters except the intercepts, and with price as a factor variable.

Comparing the coefficients in Table 4 with those of the multinomial logistic model, we find that with the mixed MNL model, the magnitude of the coefficients is slightly higher, with the exception of coffee, whose value is lower and still not significant.

Some measures of variation of the coefficients at the individual level are provided by the values of the distribution of the random coefficients in Table 5. By looking at the signs of the first and third quartile, it shows that a portion of 100g for the first course is always the preferred option with respect to 60g, but not as much as one of 80g (the value of the third quartile is positive). In general, larger portions are overall preferred with respect to smaller ones for the first and second course and for the side dish. The same holds for the inclusion of fruit and dessert. The range of price of €2.80 is mainly preferred than that of €2.50, while the ranges of €3.10, €3.40 and €3.70 are avoided as much as possible.

	Estimate	Error	p-value
first.course60g	-1.5348250	0.1905002	**8.882e-16**
first.course80g	-0.2703450	0.1357082	**0.0463596**
second.course60g	-1.0949532	0.1695011	**1.048e-10**
second.course80g	-0.4050596	0.1445987	**0.0050902**
side.dish60g	-0.6280994	0.1626479	**0.0001126**
side.dish80g	-0.1818149	0.1527812	0.2340330
fruityes	0.6312104	0.1268062	**6.433e-07**
dessertyes	0.6427777	0.1392631	**3.920e-06**
coffeeyes	0.0055156	0.1264925	0.9652199
price2.8	-0.2930953	0.1764217	0.0966464
price3.1	-0.6669026	0.1859096	**0.0003342**
price3.4	-1.8623710	0.2402937	**9.104e-15**
price3.7	-2.8937363	0.3078019	$<$ **2.2e-1**

Table 4. Coefficients of the mixed MNL model.

	1st quartile	Median	3rd quartile
first.course60g	-2.1066903	-1.534824963	-0.962959588
first.course80g	-0.6864343	-0.270344967	0.145744374
second.course60g	-1.9069572	-1.094953187	-0.282949200
second.course80g	-0.5628964	-0.405059607	-0.247222793
side.dish60g	-0.8820248	-0.628099382	-0.374173986
side.dish80g	-0.3119387	-0.181814882	-0.051691021
fruityes	0.2942503	0.631210398	0.968170535
dessertyes	0.3284038	0.642777652	0.957151516
coffeeyes	-0.7854576	0.005515604	0.796488835
price2.8	-0.5893300	-0.293095344	0.003139356
price3.1	-0.8236005	-0.666902647	-0.510204753
price3.4	-2.4198664	-1.862371045	-1.304875663
price3.7	-4.0488418	-2.893736307	-1.738630859

Table 5. Distributions of the random coefficients of the mixed MNL model.

It is interesting that the distribution of the coffee variable that has approximately the same value at the first and third quartile, but with inverted signs. This, in addition to the not statistically significance of the average coefficient at the aggregated level can be due to the fact that individuals who prefer coffee are as many as those that did not considered it important in the choice of the best *Pasto Lesto* profile. The representation of the fitted values at the individual level in Figure 5 shows that the distribution, even if slightly tailed, is almost centered on the mean value of 0.

To test if even in this case the model with price as factor had a better goodness of fit to the data, we estimated the same mixed MNL model with price as numeric and we per-

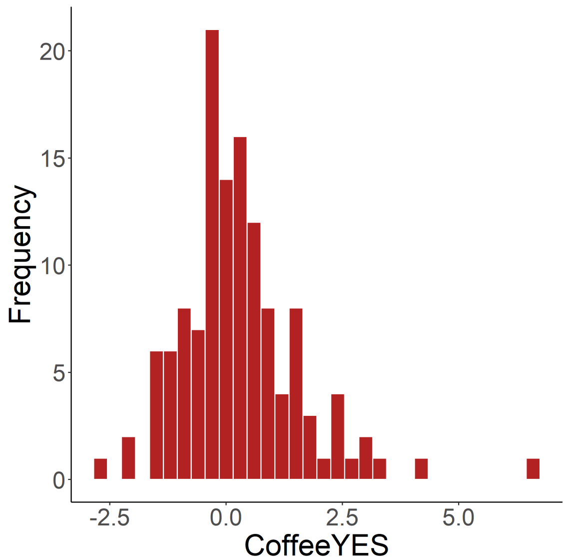

Figure 5. Histogram of the individual part worths related to coffee.

formed the Likelihood Ratio test to compare the two models. The coefficients remained the same as those in Table 4, with price having a statistically significant average coefficient of -2.57. The significance of the p-value (0.053) of the Likelihood Ratio test lead us to confidently reject the hypothesis of the two models being equal and made us consider the model with price as numeric and with a larger value relative to the logLikelihood as better.

With analogy to the MNL model, the computation of the willingness to pay leads almost to the same results: students are willing to pay €0.61 more to move from a portion of the first course of 60g to one of 100g, €0.40 more from 60g of the second course for 100g of the same course and so on. Again, the willingness to pay is useful to validate the importance of the single components of the *Pasto Lesto* attributed by individuals, as it conveys the same information as the output of the model: higher preferences for larger portions, inclusion of a fruit and a dessert and low price.

To better assess how the individual preferences vary across students and how they are related to each other, it can be useful to evaluate the mixed MNL model allowing the random coefficients to be correlated to see if students preferring some attributes also tend to prefer specific others. Table 6 provides the correlations among the random coefficients with the statistically significant values in evidence.

The statistical significant correlations are found among first course of 60g and first course of 80g (0.994), first course of 60g and side dish of 60g (0.589), first course of 80g and side dish of 60g (0.632), side dish of 60g and fruit (0.577), side dish of 60g and dessert (0.517), side dish of 80g and dessert (0.755), and first course of 60g and coffee (0.331). For those values that are ≥ 0.5 we can con-

	1st60g	1st80g	2nd60g	2nd80g	side60g	side80g	fruitY	dessertY	coffeeY	price
1st60g	1.000	**0.994**	-0.399	-0.056	**0.589**	-0.075	0.092	-0.073	**0.331**	-0.171
1st80g	0.99	1.000	-0.326	0.045	**0.632**	-0.062	0.090	-0.015	0.323	-0.139
2nd60g	-0.399	-0.326	1.000	0.741	0.039	-0.107	0.032	0.365	0.293	0.009
2nd80g	-0.056	0.045	0.741	1.000	0.356	0.052	-0.090	0.522	-0.026	0.263
side60g	0.589	0.632	0.039	0.356	1.000	0.324	**0.577**	**0.517**	-0.039	-0.024
side80g	-0.075	-0.062	-0.107	0.052	0.324	1.000	0.209	**0.755**	-0.262	0.060
fruitY	0.092	0.090	0.032	-0.090	0.577	0.209	1.000	0.102	-0.092	0.200
dessertY	-0.073	-0.015	0.365	0.522	0.517	0.755	0.102	1.000	-0.046	-0.207
coffeeY	0.331	0.323	0.293	-0.026	-0.039	-0.262	-0.092	-0.046	1.000	-0.553
price	-0.171	-0.139	0.009	0.263	-0.024	0.060	0.200	-0.207	-0.553	1.000

Table 6. Correlation coefficients of the random coefficients of the mixed MNL model.

clude that there is a not indifferent correlation between their choice by respondents.

Evaluating the models estimated so far with the Likelihood Ratio test we found that the the mixed multinomial logistic model with correlations among random coefficients has the best goodness of fit to the data.

6. Conclusions

The present research on the *Pasto Lesto* meal option offered by the food service of the canteens of the University of Trento aims at analysing students' preferences towards the portions of the components and the presence of extras in the meal in order to increase students satisfaction and improve the offered service. This was done with choice-base conjoint analysis, a technique that, with multinomial logistic models and mixed multinomial logistic models, allows to directly investigate the propensity of individuals to choose the attribute profile that maximises their utility within a limited set of alternatives.

The data required for the application of the choice-based conjoint analysis were collected through a *Telegram* chatbot that in a conversational manner provided respondents 7 questions relative to the *Pasto Lesto* and 3 questions about individual characteristics.

The questions relative to the meal were made of three attributes profiles among which the respondent had to choose the preferred one. The attribute profiles were created with a completely randomized design, by randomly sampling with replacement for each question 3 profiles from a choice set made of all the possible combinations of the levels of the *Pasto Lesto* attributes.

The collection of the data for two weeks provided us complete surveys of 127 students. After performing some exploratory analysis, we found out that there is a balance between females and males in our sample and that the larger part of the respondents eat at the university canteens at least once a week and go back to their residential home every week. Also, exploring the relationship between canteen attendance and going home frequency, we found that the majority of those that eat every day in the canteen go home every week and every month.

At the aggregated level, the best fit of the data with the the multinomial logistic model was with all the variables as factor and without the intercepts. This was used to get the estimates of the propensity of choice of each *Pasto Lesto* attribute. We found that the most important components of the meal are the first course 100g, the second course 100g, the side dish 100g, the inclusion of fruit and dessert and the price of €2.50.

Considering the price variable as numeric allowed us to interpret the other variables in monetary terms through the formulation of the willingness to pay. We found that respondents were willing to pay around €0.63 more to increase their first course from 60g to 100g, and a decreasing monetary amount for the other variables.

Then, by imaging to propose our own *Pasto Lesto* profile and to compare it to other alternatives, we simulated a preference shares of different combinations in order to see which was preferred by respondents. In this case, students seemed to prefer a discrete balance of first course, second course and side dish with possibly the inclusion of some extras at a reasonable low price (€2.50 or €2.80).

In the end, the mixed multinomial logistic model was introduced to allow the presence of heterogeneity among the preferences of the respondents and to assess how the estimates of the choice coefficients vary across students. In this case the best model used price as numeric variable, had no intercepts and allowed the random coefficients to have correlations among each other. The estimated coefficients were almost the same as those of the MNL model, but their distribution showed some heterogeneity in the preference of the first course 80g with respect to 100g and in the inclusion of fruit, dessert and coffee. In particular, a further investigation in the individual part worth of the last one showed that respondents are almost equally divided into those that prefer coffee and those that do not. This is the reason why its coefficients is near 0 and not statistically significant.

In the end, we found the presence of some statistically significant correlations among some attributes choice, even if the motivation of this is not clearly evident. Also, we tried to include some interactions with the individual variable without obtaining relevant results.

6.1. Future work and improvements

At the end of our analysis there are a number of things that we think could be interesting to change or to improve. For example a straightforward way to improve the number of respondents could be to use also other platform for the delivery of our survey (e.g. implementing chatbots for different instant messaging services).

Another interesting area is about the method with which the different profiles in each questions are generated for the user, as different techniques have different pros and cons. In a possible reiteration of the survey it can be interesting to change the individual specific questions to be asked, and see if it is possible to find some characteristic more related to the meal's preferences.

Also it could be interesting to study if and how the preferences of the students change with time, and using the tools provided by conjoint analysis if it is possible to make a more intelligent planning of the menu.

7. Bibliography

Croissant, Y., *Estimation of multinomial logit models in R : The mlogit Packages*. URL `http://www2.uaem.mx/r-mirror/web/packages/mlogit/vignettes/mlogit.pdf`

Rao, V., (2013), *Applied Conjoint Analysis*, 10.1007/978-3-540-87753-0.

Train, K., Croissant, Y., *Kenneth Train's exercises using the mlogit package for R*. URL `https://cran.r-project.org/web/packages/mlogit/vignettes/e1mlogit.html`

A. Tools used

The choice of using a telegram chatbot allowed us to avoid expensive services, but required a solution to host our service, possibly a free one.

In particular, we decided to develop this chatbot using the python API for Telegram, in particular the *Python-Telegram-Bot* Wrapper. This allowed us to quickly develop and experiment different configurations for our bot given the good documentation of the package and the community.

In theory, the script could have run on one of our computers, but it required an always-on system so this was not feasible. In order to start quickly the data collection, we decided to run the chatbot code on a *Raspberry Pi 4 model B*, the script did not require much computational power, so it ran smoothly on this system usually used of IoT projects, allowing us to use a less energy consuming machine.

The problem with this solution was that this setting was not really scalable, there was the problem of backups and stable energy supply for example. For the project at study this was not really a relevant issue, but with scalability and possible future developments in our minds we decided to make a step further.

Our aim was to make our chatbot completely online, for this purpose we decided to use *Heroku*, a platform for hosting web apps that has a free tier which is not really powerful but it is reasonable for our application and to use *MongoDB* as our no-SQL database, which also has a free tier version with limited storage capabilities which again are enough for the task.

The transition was smoother than expected because with *Docker* the process was really similar to the first one of deploying the app on the *Rapsberry*, and because *MongoDB* accepts JSON formatted data as input, which was also the format we used in the first version of our service.