Understanding Consumer Preferences for Vacation Packages: A Choice-Based Conjoint Analysis

Andrea Bortoluzzi (andrea.bortoluzzi-1@studenti.unitn.it)
Ricardo Esquivel D'Avanzo (r.esquiveldavanzo@studenti.unitn.it)
Joaquin Lopez Calvo (joaquin.lopezcalvo@studenti.unitn.it)
Maria Starodubtseva (mariia.starodubtseva@studenti.unitn.it)

1. Introduction

Understanding consumer preferences is crucial for designing products and services that resonate with target audiences. Conjoint analysis is a widely used method for modeling decision-making processes, enabling researchers to quantify the importance of different product attributes and the trade-offs consumers are willing to make. By presenting respondents with hypothetical scenarios, this approach allows for a detailed exploration of how specific features influence consumption choices.

This study focuses on vacation planning, a domain where consumer decisions are influenced by a range of factors, including price, duration, accommodation, transport, and destination. The goal of the analysis is to identify the key drivers of consumer preferences for vacation packages and to provide actionable insights for the travel and tourism industry. These findings can help businesses design offerings that better meet the needs and expectations of potential customers. The github repository for this project is available here.

2. Data Description and Methodology

The dataset for this study was created synthetically and is specifically designed to simulate consumer decision-making in vacation planning. It is organized in a long format, a standard structure for Choice-Based Conjoint (CBC) analysis, where each row represents an alternative, relative to a specific question answered by a specific respondent. In our case:

- **Respondent ID (`resp.id`)**: Identifies each participant.

- **Choice Task ID (`ques`)**: Identifies each decision-making scenario (choice task) presented to respondents.
- **Alternative ID (`alt`)**: Represents each vacation package alternative within a choice task.
- **Attributes and Levels**:
  - **Price**: Cost of the package, measured in euros (`500`, `800`, `1500`, `2500`).
  - **Duration**: Vacation length, measured in nights (`2`, `5`, `10`).
  - **Accommodation**: Type of lodging (`budget`, `3stars`, `5stars`).
  - **Transport**: Mode of transport (`bus`, `train`, `plane`).
  - **Destination**: Vacation destination (`Greece`, `Turkey`, `Portugal`, `Spain`).
- **Choice Indicator (`choice`)**: A binary variable indicating whether a given alternative within a choice task was selected by the respondent (`1`) or not (`0`). This variable is the dependent variable in the CBC analysis.

Each respondent faced multiple choice tasks, with each task consisting of four vacation package alternatives. In each task, only one alternative is marked as selected (1), and the others are marked as unselected (0). This structure is ideal for CBC analysis, since it models decision-making processes by analyzing the choices made across different tasks.

To illustrate the structure of the dataset, let us examine the first few observations (see Table 1). In the first four rows, respondent 1 selects the third alternative (price €500, duration 5 days, accommodation 5 stars, transport by bus, and destination Turkey) when faced to question 1, as indicated by a `choice` value of 1. The other alternatives were not chosen, as reflected by a `choice` value of 0.

Table 1: Sample of the Dataset

resp.id	ques	alt	Price	Duration	Accommodation	Transport	Destination	choice
1	1	1	500	10	budget	train	Greece	0
1	1	2	500	10	budget	plane	Greece	0
1	1	3	500	5	5stars	bus	Turkey	1
1	1	4	2500	2	5stars	train	Spain	0
1	2	1	800	10	5stars	plane	Greece	0
1	2	2	1500	10	5stars	plane	Turkey	0
1	2	3	500	5	3stars	plane	Spain	0
1	2	4	800	10	5stars	plane	Turkey	1

3. Data Exploration

The dataset contains 250 respondents who completed 18 choice tasks each, resulting in 72 observations per respondent and a total of 18000 observations.

Before proceeding with the analysis, we will conduct an exploratory data analysis to gain insights into the distribution of variables, identify potential issues, and prepare the data for modeling.

Table 2: Data summary

Name	vacation
Number of rows	18000
Number of columns	9
Column type frequency:	
character	3
numeric	6
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Accommodation	0	1	6	6	0	3	0
Transport	0	1	3	5	0	3	0
Destination	0	1	5	8	0	4	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
resp.id	0	1	125.50	72.17	1	63.00	125.5	188.00	250	
ques	0	1	2250.50	1299.07	1	1125.75	2250.5	3375.25	4500	
alt	0	1	2.50	1.12	1	1.75	2.5	3.25	4	
Price	0	1	1323.19	769.20	500	500.00	800.0	1500.00	2500	
Duration	0	1	5.65	3.30	2	2.00	5.0	10.00	10	
choice	0	1	0.25	0.43	0	0.00	0.0	0.25	1	

Upon reviewing the summary statistics, it is observed that there are no missing values in the dataset (see Table 2). However, some variables have incorrect data types, which are corrected to ensure proper analysis. Specifically, the `Price`, `Duration`, `Accommodation`, `Transport`, and `Destination` variables were converted into factors. This transformation ensures that each categorical attribute is correctly handled in the CBC analysis.

The distribution of the response variable indicates that the choices are relatively balanced across the four alternatives, suggesting no significant bias in the data. This balance is essential for accurate estimation of preference parameters. Additionally, the frequencies of attributes show a relatively even distribution, indicating that the levels of each attribute are well-represented in the dataset.

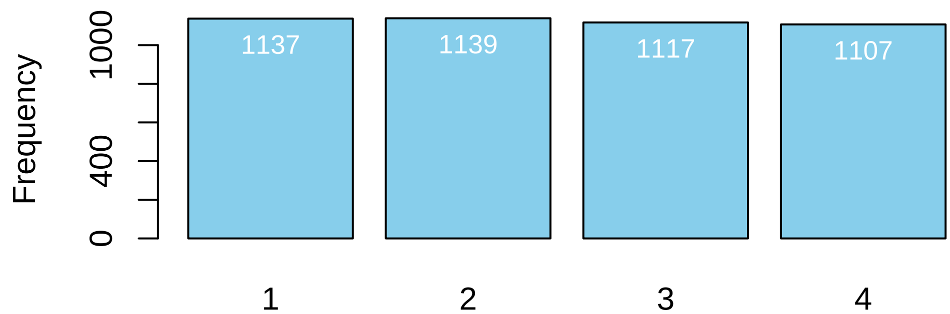


Figure 1: Distribution of choice between the alternatives

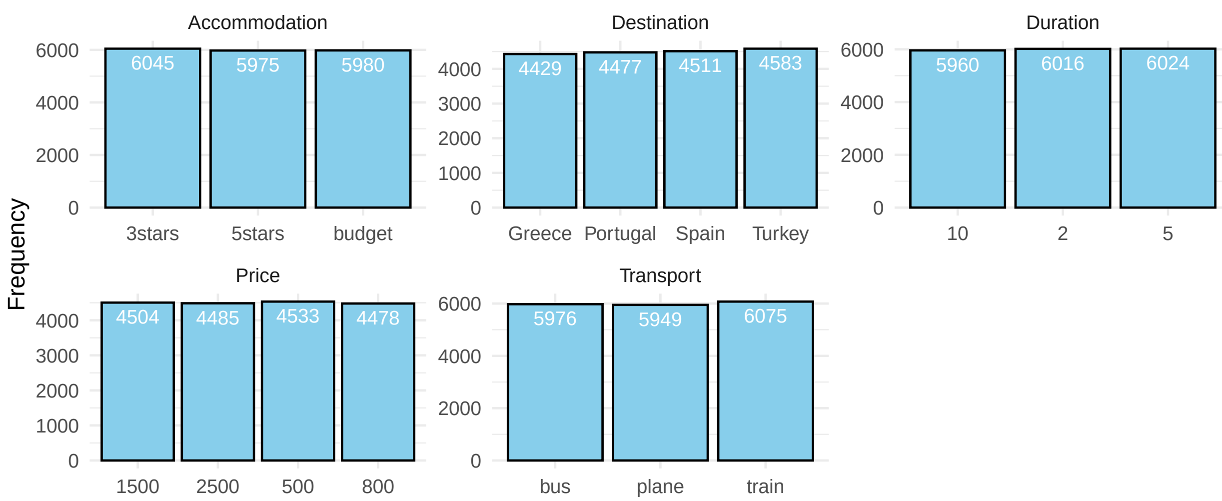


Figure 2: Distributions of attributes

4. Fitting Models with Fixed Effects

Our modeling starts from discrete choice modeling approach, specifically a Multinomial Logit (MNL) regression. This method will help us to examine the relationship between consumer choices and product attributes. This first model assumes homogeneous preferences across respondents by estimating average effects for each attribute level.

Let us first fit a baseline MNL model to the data with the following specification:

- Dependent Variable: `choice`
- Independent Variables: `Price`, `Duration`, `Accommodation`, `Transport`, and `Destination`
- Model Characteristics: This model includes **intercepts** and only accounts for **fixed effects**, assuming homogeneous preferences across respondents without considering random variation in preferences or unobserved heterogeneity.

```
                     Estimate Std. Error   z-value Pr(>|z|)
(Intercept):2        -0.02650    0.04769  -0.55566  0.57844
(Intercept):3        -0.01744    0.04795  -0.36374  0.71605
(Intercept):4        -0.05441    0.04803  -1.13288  0.25726
Price800             -0.40834    0.05231  -7.80556  0.00000
Price1500            -0.78829    0.05434 -14.50756  0.00000
Price2500            -1.21432    0.05831 -20.82503  0.00000
Duration5            -1.13564    0.05002 -22.70245  0.00000
Duration10           -0.76635    0.04729 -16.20525  0.00000
Accommodation3stars   0.32385    0.05123   6.32195  0.00000
Accommodation5stars   1.29993    0.05037  25.80621  0.00000
Transportplane        0.49622    0.04719  10.51546  0.00000
Transporttrain       -0.22285    0.04944  -4.50750  0.00001
DestinationPortugal   1.64963    0.06075  27.15520  0.00000
DestinationSpain      1.13903    0.06027  18.89880  0.00000
DestinationTurkey     0.46241    0.06270   7.37545  0.00000
```

The alternative-specific constants represent preferences for the positions of the alternatives within each question. Specifically, the parameters `(Intercept):2`, `(Intercept):3`, and `(Intercept):4` quantify the preference for each position from left to right relative to the first position on the left. The estimated MNL model indicates that these intercepts are very small and not statistically significant. To improve model parsimony and precision, we have chosen to exclude these intercepts from the analysis, thereby reducing the degrees of freedom.

We formally evaluate the decision to remove the intercepts using a likelihood ratio test. This test compares the larger model `model1`, which includes the intercepts, with a smaller `model2` that excludes them. The comparison between the full model and the reduced model without intercepts yields a p-value of `0.7185`. Given this high p-value, we conclude that there is no significant difference in goodness of fit between the two models, indicating that both explain the data equally well. Consequently, the alternative-specific constants are not essential for adequately modeling the data.

```
Likelihood ratio test

Model 1: choice ~ Price + Duration + Accommodation + Transport + Destination |
    -1
Model 2: choice ~ Price + Duration + Accommodation + Transport + Destination
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  12 -4798.1
2  15 -4797.4  3 1.3451     0.7185
```

We continue to explore opportunities to simplify the model. In the previous analysis, the attribute `price` was treated as a qualitative variable. We now incorporate it as a quantitative predictor in order to understand if the model's predictive power remains consistent.

```
Likelihood ratio test

Model 1: choice ~ as.numeric(as.character(Price)) + Duration + Accommodation +
    Transport + Destination | -1
Model 2: choice ~ Price + Duration + Accommodation + Transport + Destination |
    -1
  #Df  LogLik Df  Chisq  Pr(>Chisq)
1  10 -4811.9
2  12 -4798.1  2 27.512 0.000001061 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Again, we formally evaluate the decision to transform the attribute from qualitative, to quantitative, using a likelihood ratio test. Given the extremely low p-value, we reject the null hypothesis and select `model2` over `model3`. That is, the model with price as a qualitative variable is superior in terms of goodness of fit.

Interpretation of the Selected Model

The selected model (`model2`) provides valuable insights into consumer preferences for vacation packages. On average, we observe a clear price sensitivity, as the prices increase, the likelihood of selection decreases. Regarding the duration of the offered options, the most favored option is a short 2 night trip, followed by a longer 10 night trip. The intermediate 5 day trip is the least preferred. Accommodation preferences show a strong inclination towards 5-star hotels, followed by 3-star. Low cost options are ranked lowest. For transportation options, planes are most preferred, followed by bus and trains being the least desirable. Regarding the destination, the most preferred country is Portugal, followed by Spain and then Turkey. Greece is the least desirable destination on average.

	Estimate	Std. Error	z-value	Pr(>|z|)
Price800	-0.40830	0.05229	-7.80885	0.00000
Price1500	-0.78844	0.05432	-14.51609	0.00000
Price2500	-1.21408	0.05830	-20.82355	0.00000
Duration5	-1.13604	0.05002	-22.71352	0.00000
Duration10	-0.76658	0.04728	-16.21255	0.00000
Accommodation3stars	0.32398	0.05122	6.32505	0.00000
Accommodation5stars	1.29937	0.05036	25.79909	0.00000
Transportplane	0.49577	0.04716	10.51151	0.00000
Transporttrain	-0.22399	0.04942	-4.53238	0.00001
DestinationPortugal	1.64899	0.06074	27.14913	0.00000
DestinationSpain	1.13827	0.06025	18.89151	0.00000
DestinationTurkey	0.46206	0.06269	7.37042	0.00000

Willingness to Pay

We can further improve our interpretation of the model by calculating the willingness to pay for each level's attribute utilizing the version of the model that treats price as a quantitative attribute. We obtain this value computing the ratio between the model parameter for that level and the price parameter. The results represent the monetary value that respondents are willing to pay for changes in each attribute.

```
as.numeric(as.character(Price))                       Duration5
                        -1.0000                      -1948.5319
                     Duration10             Accommodation3stars
                     -1312.8098                        555.5628
            Accommodation5stars                  Transportplane
                      2219.0904                        848.0462
                 Transporttrain             DestinationPortugal
                      -386.1155                       2817.9713
               DestinationSpain               DestinationTurkey
                      1947.5294                        790.8334
```

Upon analysis of the results, we find strong preferences on the monetary value respondents place on various vacation attributes. For trip duration, respondents are willing to pay up to €1948.53 more for a 2-night vacation instead of a 5-night vacation and €1312.81 more for a 2-night vacation instead of a 10-night vacation, showing a strong preference for shorter trips. Regarding accommodation, customers value high-end options significantly, being willing to pay €555.56 more for a 3-star hotel over a budget hotel and €2219.09 more for a 5-star hotel over a budget option. In terms of transportation, respondents are willing to pay €848.05 more to fly instead of taking a bus, while preferring a bus over a train by €386.12. Finally, regarding destination preferences, an average customer is willing to pay €2817.97 more to visit Portugal over Greece, €1947.53 more to visit Spain over Greece, and €790.83 more to visit Turkey over Greece.

In summary, on average respondents prefer 2-night vacations, 5-star hotels, plane travel, and destinations like Portugal and Spain, showing a clear willingness to pay more for convenience and luxury.

5. Fitting Models with Random Effects

In this section, we fit mixed logit models in order to capture individual differences in preferences, addressing limitations of the standard MNL model. By including random coefficients for attributes, the mixed logit accounts for varying preferences across respondents. We later compare the models to the MNL model (`model2`) using a likelihood ratio test to assess improvements in model fit.

Uncorrelated Random Coefficients

First, let us fit a mixed MNL model without correlation of random coefficients. The results show the estimated standard deviations for each attribute, which represent the level of heterogeneity in preferences across respondents.

```
                       Estimate Std. Error   z-value Pr(>|z|)
Price800               -0.46671    0.05871  -7.94987  0.00000
Price1500              -0.89841    0.06097 -14.73590  0.00000
Price2500              -1.40740    0.06758 -20.82495  0.00000
Duration5              -1.31978    0.05919 -22.29836  0.00000
Duration10             -0.89200    0.05445 -16.38098  0.00000
Accommodation3stars     0.38422    0.05666   6.78157  0.00000
Accommodation5stars     1.42876    0.05869  24.34366  0.00000
Transportplane          0.53792    0.05262  10.22311  0.00000
Transporttrain         -0.30868    0.05695  -5.41968  0.00000
DestinationPortugal     1.86934    0.06942  26.92851  0.00000
DestinationSpain        1.27227    0.06644  19.14997  0.00000
DestinationTurkey       0.47413    0.06887   6.88462  0.00000
sd.Price800            -0.28235    0.08646  -3.26580  0.00109
sd.Price1500            0.42211    0.08726   4.83717  0.00000
sd.Price2500            0.28194    0.09560   2.94931  0.00318
sd.Duration5            0.68053    0.07578   8.98045  0.00000
sd.Duration10           0.70006    0.07332   9.54827  0.00000
sd.Accommodation3stars -0.40457    0.07684  -5.26517  0.00000
```

```
sd.Accommodation5stars  0.62779    0.06408   9.79669  0.00000
sd.Transportplane       0.45004    0.07278   6.18347  0.00000
sd.Transporttrain       0.82148    0.06687  12.28541  0.00000
sd.DestinationPortugal -0.11650    0.08256  -1.41120  0.15819
sd.DestinationSpain    -0.07090    0.08846  -0.80152  0.42283
sd.DestinationTurkey    0.57844    0.07811   7.40541  0.00000
```

We notice that almost all the standard deviation coefficients are statistically significant with low p-values. The biggest variability is observed for transportation by train, longer stays, 5-star accommodation and Turkey as a destination. This indicates that respondents have diverse preferences for these attributes, with some individuals valuing them more than others. No variability is observed for the Portugal and Spain as destinations, suggesting that respondents have similar preferences for these countries.

The likelihood ratio test also indicates that the inclusion of random coefficients improves the model's fit significantly, reinforcing the importance of accounting for individual differences in choice behavior.

```
Likelihood ratio test

Model 1: choice ~ Price + Duration + Accommodation + Transport + Destination |
    -1
Model 2: choice ~ Price + Duration + Accommodation + Transport + Destination |
    -1
  #Df  LogLik Df  Chisq            Pr(>Chisq)
1  12 -4798.1
2  24 -4605.0 12 386.16 < 0.00000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Correlated Random Coefficients

To better understand the associations among random coefficients, we fit a mixed MNL model with correlated random coefficients, allowing for the estimation of a full covariance matrix to capture interdependencies among the random effects.

```
Likelihood ratio test

Model 1: choice ~ Price + Duration + Accommodation + Transport + Destination |
    -1
Model 2: choice ~ Price + Duration + Accommodation + Transport + Destination |
    -1
  #Df  LogLik Df  Chisq            Pr(>Chisq)
```

```
1  24 -4605.0
2  90 -4462.4 66 285.22 < 0.00000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The extremely low p-value (< 0.001) of the likelihood ratio test suggests that the additional complexity in Model 2 (model2.mixed2) significantly improves the model's ability to explain the data compared to Model 1 (model2.mixed). Therefore, we select the mixed MNL model with correlated random coefficients as the best model so far to capture individual differences in preferences.

Strongly Correlated Features

Finally, we try fitting a model with only the strongly correlated features (correlation ≥ 0.7) to check if the model may be simplified without losing much information. This can help reduce the complexity of the model and improve its interpretability.

```
Likelihood ratio test

Model 1: choice ~ Price + Duration + Accommodation + Transport + Destination |
    -1
Model 2: choice ~ Price + Duration + Accommodation + Transport + Destination |
    -1
  #Df  LogLik Df  Chisq      Pr(>Chisq)
1  52 -4521.0
2  90 -4462.4 38 117.11 0.0000000005423 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The likelihood ratio test indicates that the model with all correlated random effects (model2.mixed2) is superior to the model with only strongly correlated features (model2.mixed3). Therefore, we conclude that the mixed MNL model with all correlated random coefficients is the best model to capture individual differences in preferences.

Interpretation of the Best Model

The best mixed MNL model (`model2.mixed2`) provides insights into heterogeneity in preferences among travelers by estimating random coefficients and their correlations.

	Estimate	Std_Error	z_value	p_value
sd.Price1500	0.46661	0.07728	6.03773	0.00000
sd.Price2500	0.49906	0.07680	6.49811	0.00000
sd.Duration5	0.73210	0.07380	9.91977	0.00000
sd.Duration10	0.77400	0.07515	10.29912	0.00000
sd.Accommodation3stars	0.31755	0.07223	4.39623	0.00001
sd.Accommodation5stars	0.92116	0.07591	12.13459	0.00000
sd.Transportplane	0.79266	0.07930	9.99613	0.00000
sd.Transporttrain	1.07486	0.07950	13.52024	0.00000
sd.DestinationPortugal	0.83207	0.09438	8.81640	0.00000
sd.DestinationSpain	0.82288	0.09561	8.60641	0.00000
sd.DestinationTurkey	1.03866	0.10054	10.33050	0.00000
cor.Price1500:Price2500	0.44222	0.18853	2.34562	0.01900
cor.Price2500:Duration10	-0.88194	0.32897	-2.68096	0.00734
cor.Accommodation5stars:Transporttrain	-0.35970	0.17929	-2.00625	0.04483
cor.Transportplane:Transporttrain	0.36717	0.09241	3.97306	0.00007
cor.DestinationPortugal:DestinationSpain	0.77721	0.06710	11.58321	0.00000
cor.DestinationPortugal:DestinationTurkey	0.48061	0.10057	4.77867	0.00000
cor.DestinationSpain:DestinationTurkey	0.80215	0.06643	12.07578	0.00000

The output above contains only the estimates for the standard deviations and correlations with statistically significant p-values (< 0.05), illustrating the most relevant heterogeneity in preferences among respondents and associations among attributes.

Once again, the results show that the standard deviations of random coefficients are statistically significant almost for all the attributes, indicating substantial heterogeneity in preferences among respondents. Interestingly, there exists some variability in the higher price range, with the `Price1500` and `Price2500` coefficients having significant standard deviations, but low variability in the lower price ranges (insignificant p-values are not shown in the table above).

In addition, the correlations between random coefficients reveal interesting patterns of association among attributes. First, there is a high correlation between the following pairs of destinations: `Portugal-Spain`, `Portugal-Turkey`, `Spain-Turkey`. Primarily, this means that if a consumer prefers one of these options, they are likely to also enjoy the other in the pair. Additionally, offers that combine the destinations should be well received. Second, the positive correlation between `Price1500` and `Price2500` indicates that individuals who prefer/refuse higher-priced options might also prefer/refuse even more expensive alternatives. Moreover, there is a moderate correlation between `Transportplane` and `Transporttrain`, suggesting some alignment in preferences for these two modes of transportation.

The analysis of the distributions of random coefficients in the best mixed MNL model further reveal several relevant preferences among respondents (see Figure 3-4). First, while the majority of individuals prefer shorter stays, approximately 9% actually favor 10-day stays over 2-day ones. Regarding accommodation, 8% of respondents prefer budget options over 3-star accommodations, yet 96% of respondents prefer 5-star hotels over budget options.

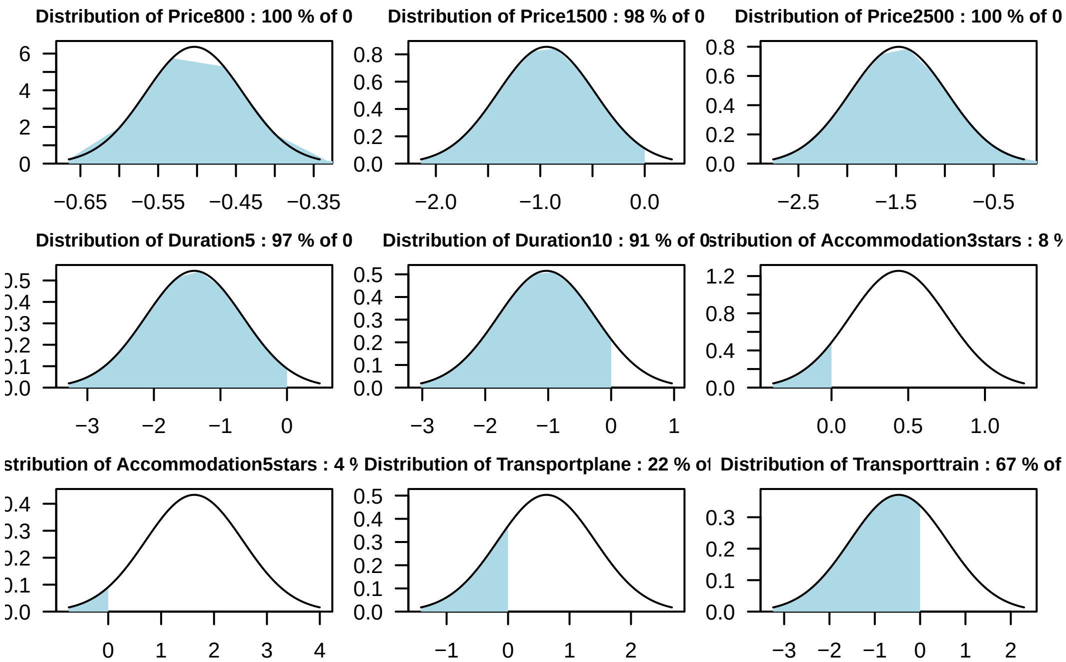


Figure 3: Distributions of Random Coefficients in the Best Mixed MNL Model

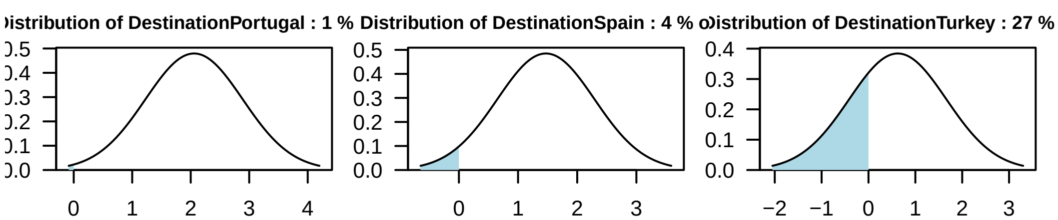


Figure 4: Distributions of Random Coefficients in the Best Mixed MNL Model

In terms of transportation, 22% of respondents prefer traveling by bus over flying, and 67% prefer the bus to the train. When considering destinations, although almost everyone prefers Portugal and Spain over Greece, 27% of respondents actually prefer Greece to Turkey. Finally, the analysis shows a clear preference for cheaper options, as respondents consistently favor lower-priced choices.

6. Preference Shares and Sensitivity Chart

Preference Shares

Another useful approach to assess the role of product attributes consists on using the fitted model to obtain preference share predictions. By varying the attribute levels of the planned design, we can observe how variations in the design influence the preference shares.

The reference package includes the following attributes:

- **Price**: 800,
- **Duration**: 2 days,
- **Accommodation**: 5 stars,
- **Transport**: Plane,
- **Destination**: Portugal.

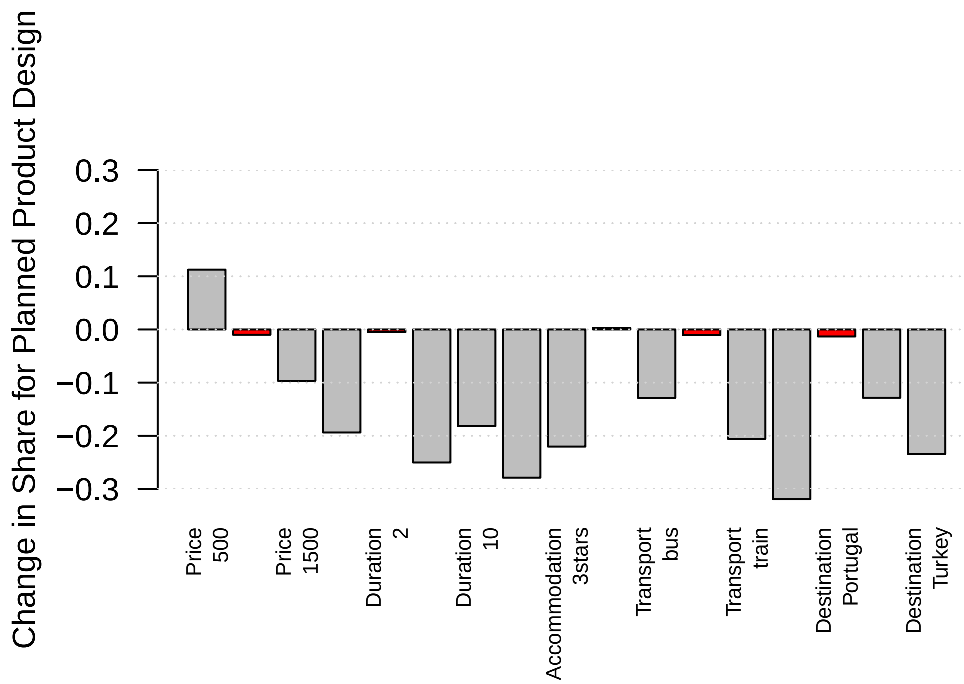
This package was chosen as the reference based on results from the models fitted earlier.

	colMeans(shares)	Price	Duration	Accommodation	Transport	Destination
1	0.425513832	800	2	5stars	plane	Portugal
2	0.236316197	500	2	5stars	bus	Spain
3	0.094806478	800	2	budget	plane	Portugal
4	0.040901840	500	2	3stars	plane	Greece
5	0.035824293	800	2	5stars	bus	Greece
6	0.033784535	2500	10	5stars	plane	Spain
7	0.021960947	1500	10	5stars	plane	Greece
8	0.018912620	500	2	3stars	train	Greece
9	0.016917153	1500	10	5stars	train	Turkey
10	0.015309410	800	10	budget	train	Spain
11	0.012832223	800	5	budget	plane	Turkey
12	0.012796634	1500	10	3stars	bus	Spain
13	0.010562728	1500	5	budget	bus	Portugal
14	0.010324842	2500	5	3stars	train	Portugal
15	0.010100445	500	5	budget	train	Turkey
16	0.003135822	1500	5	budget	bus	Turkey

The table above shows the preference shares for the reference package and the 15 alternative designs. The reference package has the highest preference share of 0.43 (PS = 0.43), followed by the design with the destination in Spain, 5-star accommodation, bus transportation, 2-day duration, and price of 500 euros (PS = 0.24). The alternative design with the lowest preference shares is the one with the destination in Turkey, budget accommodation, bus transportation, 5-day duration, and price of 1500 euros (PS = 0.003).

Sensitivity Chart

Sensitivity chart is yet another useful tool to visualize how changes in attribute levels affect preference shares. The sensitivity chart below shows the impact of varying attribute levels on the preference shares of the base package (colored in red).



Portugal stands out as the most attractive destination, while Spain, reducing the preference shares only by 12%, remains somewhat competitive. In contrast, Turkey and Greece reduce the shares by larger margins (24% and 30%, respectively) making them much less preferred destinations.

In terms of transportation, traveling by plane is the most preferred option. Bus travel reduces the share by around 12%, reflecting moderate aversion, while train travel decreases preference share by almost 20%, indicating it is the least preferred mode of transport.

For accommodations, 3-star accommodations reduce preference share by 22%, suggesting that mid-tier options are less competitive. Budget accommodations decrease preference share by almost 30%, indicating they are highly unattractive compared to 5-star accommodations.

Regarding trip duration, a 2-day vacation is the most appealing choice, followed by a 10-day trip which reduces preference share by 18%, indicating it is a possible alternative. A 5-day trip decreases the share by 25%, making it the least attractive option.

Finally, price has a significant influence. A price of 500 can increase the preference shares by 11%, indicating that lower pricing strongly enhances attractiveness. A mid-range price of 1500 reduces preference share by 8%, showing relatively small sensitivity, while the highest price option, 2500, reduces preference share by almost 20%, highlighting that higher prices significantly deter preferences.

7. Results

When we analyze the estimated average values over all respondents, we come to the following conclusions:

- **High price sensitivity**: as the price increases, the likelihood of selecting options decreases.
- **Short term stays are preferred**: 2 nights duration was the top choice by far. Still, a secondary offer for a 10-nights vacation could be a valid alternative.
- **Focus on high-quality Accommodations**: 5 star hotels were the preferred option by a large margin, with 3 stars options on second place.
- **Plane first, bus as an alternative**: respondents are willing to pay €848.05 more to fly instead of taking a bus, while preferring a bus over a train by €386.12.
- **Prioritize Iberian destinations**: With Greece as a baseline, the average respondent would be willing to pay €2817.97 more to visit Portugal and €1947.53 more to visit Spain. Turkey places third.

Then, when the Mixed Effects model was fitted, we checked a magnitude of variation between respondents that suggest the existence of non-negligible niches. We think the most relevant ones were:

- **10% of individuals preferred 10-day stays**: A small niche that might be suited for a specific offer on longer trips.
- **A low-budget niche**: Considering the high price sensitivity across the board, and a 17% of respondents who had budget-hotels as their 2nd preferred option (just behind 5-star hotels), would be reasonable to fit an offer for this subset.
- **Bus as the 2nd choice**: While only 12% of respondents prefer bus over planes, 65% do prefer it over the train. If a connection wouldn't be possible flying, the alternative would be clear.
- **Highly correlated destinations**: Not only Portugal and Spain are highly correlated choices, also visiting these destinations is highly correlated with a desire to visit Turkey. This point should be used for follow-up offers for those who already traveled with the company.

8. Managerial Implications and Conclusions

The main products of the company should be:

1. **Trip to Portugal including**: 2-night in a 5-star hotels and travel by plane.
2. **Replicate trip 1., but with Spain as a destination**.

Also, we encourage the company to save contact information of those tourists who already traveled with options 1. or 2. and establish remarketing campaigns offering:

- The alternative option. That is, if the client already went to Portugal, offer a trip to Spain. Inversely, if the client went already to Spain, offer Portugal. The strong correlation between these destinations motivates this recommendation.

3. **Short trip to Turkey**: following the rest of the characteristics from offers 1. and 2.

Then, the following alternative offers should be tried out as niche options. Some of these include:

4. **10-days trip through the Iberian Peninsula**: A 10% of respondents preferred a longer trip. This offer should maintain the transportation by plane as well as the 5-star hotels.
5. **Low-budget alternatives to offers 1. and 2.**: Considering the high price sensitivity of the average respondent and a niche of users that have budget-hotel

On offers 4. and 5., we recommend to avoid broad promotions (in opposition to offers 1. and 2., aimed for the massive public). Instead, the agency should try to identify the market segment and direct the promotion just towards the users belonging to those niches. For this segmentation we propose to perform a client clustering that finds more granular information on these niches, and only then target the promotions specifically to those who share (demographic/behavioural/geographic/psychographic) variables with those clusters.

Finally, some points that the agency tour planners could benefit from:

- If a trip from Point A to Point B cannot be covered by plane, bus is the best alternative.
- When targeting niche segments, such as budget-conscious travelers or those preferring longer trips, focus on tailored marketing and clustering methods to identify and reach these audiences effectively.
- Portugal and Spain are not only highly preferred destinations but also strongly correlated with interest in Turkey, providing an opportunity to cross-promote these destinations in remarketing campaigns.

Throughout this study we identified significant consumer preferences for vacation packages from the Choice-Based Conjoint Analysis perspective. By systematically analyzing attributes such as price, duration, accommodation, transportation, and destination, we came out with 5 actionable insights for the tour agency.

Key insights from the analysis highlight a strong sensitivity to price, a clear preference for short trips (2 nights), and a significant inclination toward 5-star accommodations and plane travel. Additionally, niche markets were identified and detailed in the results section. To ensure our recommendations reflect real-world consumer diversity, we combined findings from a fixed-effects multinomial logit model, which captures average respondent behavior, with a mixed-effects model featuring correlated random coefficients to account for individual differences.