

MASTER IN DATA SCIENCE FOR MANAGEMENT

SUBJECT : ANALYSIS OF QUESTIONNAIRE

PROF. : G.COMMIS

**ETNA WINERY SURVEY**



Students:

**GIANLUIGI MAZZAGLIA**

**ANAHITA ESFANDIARYFARD**

**BIAGIO INCARDONA**

**THAMIRES DE SOUZA OLIVEIRA**

Academic year 2021-2022

**Table of Contents**

[INTRODUCTION 2](#_Toc93574430)

[1. Questionnaire Structure 2](#_Toc93574431)

[1.1. Dataset 5](#_Toc93574432)

[2. Univariate Analysis 5](#_Toc93574433)

[3. Missing values Handling 5](#_Toc93574434)

[4.2 IRT- Item Response Theory 5](#_Toc93574435)

[4.2.1 Dataset level IRT 6](#_Toc93574436)

[4.2.2 IRT on topics 14](#_Toc93574437)

## INTRODUCTION

The present study, with reference to the wine, was carried out between December and January 2020/21. The purpose of the survey was to discover winery market trends and customer preferences, focusing on Etna wines. All data have been collected anonymously in line with the privacy policy. Our project was done using the statistical software SAS, in particular after a firstly description of the Questionnaire Structure we did the Univariate Analysis in order to understand the distribution of the answer. Then we handled the missing values and we performed Multivariate Analysis, in particular we focus on Item Response Theory, Factor Analysis and Latent Class Analysis.

The Questionnaire was organized in four sections, the first of which was dedicated to the wine preferences, expertise and consumers’experience ( drinking frequency, type of drinks / wine, winery experience). The second part is referred to the buying experience ( buying frequency, consumed quantities, consumption time, factors that influence the purchasing moment). The third part of the questionnaire allowed us to collect information about the knowledge an consumption of Etna DOC (where the respondents will give an answer just in case they heard about Etna wine). The fourth section allowed us to obtain information on the socio-economic characteristics of the respondent (age, gender, level of education, occupation).

# Questionnaire Structure

The Questionnaire is structured in 23 questions, that contains in total 42 elements.

*Section 1: Wine preferences, expertise and experience*

Question 1 : How much do you like the following drinks?

The questions contains 4 items and asks the respondent how much do they like drinks like wine, beer, soft drinks and cocktails. Every item is placed on a semantic scale from 1 (not at all) to 4 (a lot). This question can be useful to understand the drinking preference of the respondent.

Question 2: How much do you like the following kinds of wine?

The question contains 5 items from 1 to 4 adding the possibility to answer “Never tasted”. As before every item is placed on a semantic scale from 1 (not at all) to 4 (a lot). This question allow us to understand which kind of wine the respondent prefers within red, white, rose’, sparkling and sweet.

Question 3 : Have you ever tried a wine tasting experience?

This question asks to the respondent if he never tried a wine tasting experience, it’a a categorical variable with binary choice between yes or not. Answer this question can help us with the analysis in order to categoryze the typology of respondent, likely a person who is interested in things like this can have an high at least curiosity to improve his knowledge of wine.

Question 4: Have you ever visited a winery?

Also this question it’s a categorical variable with binary choice between yes or not. Like the previous question, it can be useful to understand if the respondent is used to going in a winery.

Question 5: Have you ever attended an in-depth wine course?

This question asks to the respondent if he never attended a wine course, it’s a categorical variable with binary choice yes or not. This allow us to better understand the respondent in terms of technical expertise.

Question 6: Which is your level of knowledge around wines?

It’s a closed item that asks about the level of knowledge around wines of the respondent, in particular it is an ordinal scale ordered from “None” (no knowledge) to “High” (professional knowledge level). This answer can also useful to understand if matches with the previous questions, a respondent who has followed a depth course it’s supposed to have an high level of knowledge around wines.

Section 2: Wine buying Experience

Question 7: How often do you buy wine on average in a month?

The question asks the respondent how often he buy wine on average in a month, it’s an ordinal variable and only one answer is allowed. The utility of this question can be to better understand if the respondent is a frequently user of wine.

Question 8: How many bottles of wine do you buy on average per month?

The question can be similar to the previous one, It’s an ordinal variable and only one answer is allowed. The difference to the previous question is to understand the quantity of bottles that the respondent buy monthly (he can choose in a range from 1-3 bottles, 4-6, to more than 12 bottles) and it can be useful to highlight if he is a consumer during meals or maybe he just want to keep some bottles to share with the others.

Question 9: How often do you buy wine in the following store?

This question contains 3 items and each of them is placed on a semantic scale from 1 (never) to 4 (every time) in order to understand if the respondent used to buy wine in a supermarket, in a winery or online, so we can better understand the typology of consumer.

Question 10: How relevant are the following features when you buy a wine?

This question contains 8 items and it’s asked to the respondent to answer a list of features that he consider important during the buying transaction. Every item is placed on a semantic scale from 1 (not at all relevant) to 4 (extremely relevant) and we find items like grape origin, grape variety, budget friendly, brand awarness, vintage, label detailed, attractive packaging, eventual promotions. This question could be extremely important to understand which of the following features can move on our respondent when he buy a wine and so we can realize the kind of person we are referring to.

Question 10: How much do you spend on a bottle of wine on average?

It’s a closed item that can be placed in an ordinal scale and ask to the respondent how much he is spend on average for a bottle of wine. The answer can be just one and this information can be useful to understand if an higher quality of wine is important for the respondent and how he can afford to spend on it.

Question 11: During the pandemic, did the frequency with which you buy wine change?

The question contains 1 item and asks to the respondent if during the pandemic the frequency with which you buy wine is changed. Every item is placed on a semantic scale with 5 points and a middle point 0. This question can be useful to understand if the pandemic period has changed the habits of the consumer.

Question 12: For what reason have you bought wine in the last 3 months?

The question asks to the respondent the reason why he bought wine in the last 3 months. The question is a multi select Multiple choice question and allows to choice more than one option (within Home consumption, to buy a gift, for a special event, to try a new wine) and also allow using an opened answer to insert “Other” if the option doesn’t belong to the alternative. The utility of this question allow us to define a reason why the respondent used to buy wine.

Section 3: Etna DOC wine

Question 13: Have you ever heard about Etna DOC wine before?

This question is mandatory and according to the answer that the respondent will give, the next question will be different. In particular the question asks if the respondent has never heard about Etna DOC. It is a categorical variable with binary choice between yes or not. This question can be useful to understand if the respondent knows the Etna DOC or not. This item performs a branching and who answers no is skipped to the final section.

Question 14: Have you ever bought Etna wine?

It is a closed item and it asks the respondent if he never bought Etna wine. It’s a categorical variable with 3 choice (Yes, No, I don’t know) and it is useful to understand if the respondent has never testef Etna wine.

Question 15: How much do you like Etna wines more than other wines?

This question asks to the respondent how much he likes Etna wine more than other wines. Every item is placed on a semantic scale from 1 (not at all) to 4 (a lot). This question allow us to understand if the respondent appreciates the taste of Etna wine.

Question 16: How much do you agree with the following statements about Etna wine?

This question contains 4 closed items and asks to the respondent how much he agrees with certain statements. Every item is placed on a Likert scale with 4 modes from 1 (not at all in agreement) to 4 (completely in agreement). The utility of this question is to understand respondents’ opinion about Etna wine.

Question 17: How likely are you to recommend Etna wine to your family and friends?

This question asks to the respondent how likely they would recommend Etna wine to family and friends. It is on a Likert scale with 4 modes from 1 (not likely at all) to 4 (extremely likely). This question is useful to understand if respondents consider Etna wine as one of the top quality wine and they also provide to recommend it to closers.

Section 4: Socio – demographic data

Question 18: Please enter your gender.

This is a closed item and the question asks to the respondent his gender. It is a categorical variable with 3 choices (man, woman, prefer not to say).

Question 19: Please enter your age:

This question is an open item on an interval scale and the answer is required. It asks the respondent his/her age and give us a discrete variable in a set that is finite as a lower bound, so the respondent must have an age not less than 18, while the upper bound is infinite.

Question 20: Please enter your level of education.

This question allow us to understand the level of education of the respondent. We have here an ordinal scale (from primary and Medium school to University degree). This question can be useful to understand if could be a correlation between people who have an higher level of education and the knowledge of Etna wine and if in general the choices during shopping time can be different.

Question 21: Where are you from?

This question is a closed item. It asks the respondent the place where he/she comes from. It is a categorical variable with multiple modes. The respondent can be Sicilian, Italian, from an EU country or not EU.

Question 22: What is your occupation?

It is a closed item and asks to the respondent what is her/his occupation. It is a categorical variable with multiple modes. If the occupation is not in the list the respondent can also add his own job applying an open item in the space next to the voice “Other”.

# Dataset

# Univariate Analysis

*Section 1: Wine preferences, expertise and experience*

Question 1 : How much do you like the following drinks?

This question is a scale of preference to each type of drink, because of that we divided the graphs of barplot and table of frequency in each kind of drink to have a better view of the results, we will analyse the 247 responses.

# 

# 

# The third drink to be analysed is cocktail, where we can check through the graph that almost the same amount of people chose both level 2, 3 or 4 preference in this type of drink, and the minority chose option 1.

# The second drink to be analysed is soft drinks preferences, like coca-cola, soda,etc. In this case the major part of the people that chose level 2 for their preference for this type of drink, whichis considered a low level. Just over 26% declared that they like this kind of beverage a lot.

# The first analysis is over beer preference, where 34% of the people declared that they like beer in a level 4 (a lot) and only 17% declared that the preference of beer is “not at all”. The pattern of the graph is the decrease of responses in the order from 4 to 1 in the level of preference.

# 

# In red wine observation 72% of people are accumulated on levels 1 and 2 of preference.

# The other drink preference to be analysed is rose wine which has its majority at level 2 of preference and only 14,2% at level for of preference.

# The next drink to be analysed is whine wine preference, in this case its possible to see that the majority chose the level 3 of preference in this kind of drink and almosto 70% of all observations are concentrated on levels 3 and 4.

# 

# At the analysis of sweet wine, we can see that we have its minority on 4 level and the majority of people, almost 60%, are concentrated on the first 2 levels.

# We can see that on sparkling wine preferences, we have less people on level 4 and we have a cumulative percent of 80,16% on levels 1, 2 and 3.

# 

# 

# Missing values Handling

As we observe from the univariate analysis the dataset contains missing values. Which is due to branching of “buying experience” question and hearing about “etna doc” wine question.

# 4.2 IRT- Item Response Theory

Item response theory (IRT) is a paradigm for the analysis, and scoring of tests, questionnaires, and similar instruments measuring abilities, attitudes, or other factors. IRT is a theory of testing based on the relationship between individuals’ performances on a test item and the test takers’ levels of performance on an overall measure of the ability that item was designed to measure. IRT is based on the idea that the probability of a correct/keyed response to an item is a mathematical function of person and item parameters. The first parameter affecting the item reponses of an item is the respondent’s level on the latent train that item is trying to evaluate, other factors can be the difficulty level (how much of skills/satisfaction is needed to respond in a correct/positive way), the discrimination power (how good is the item in discriminating between individuals with different levels of the latent trait) and the easiness of guessing of a certain item.

Item Response Theory is based on 4 main assumptions:

1. Monotonicity: if the trait level increases, the probability of a correct/positive reponse increases as well
2. Unidimensionality: there is one dominant latent trait that is measured and this trait is the driving force for the responses observed for each item
3. Local Independence: the responses for an item don’t affect the responses for other items

There are 3 main steps in applying an IRT analysis: model selection, estimation of the parameters and estimation of the abilities of the respondent.

Initially IRT will be applied to the whole dataset in order to have an overall look on the items’ difficulty, then it will be applied onto the subgrops of items extrapolated from the original questionnaire.

## 4.2.1 Dataset level IRT

In this first step, IRT will be applied to all the items together, except for the socio-demographic items (GENDER, AGE, EDUCATION, LOCATION, JOB) since for these items doesn’t make any sense to try studying their difficulties (in this context).

In the following SAS procedure, all the default options will be used. The Marginal Maximum Likelihood estimation method will be used to estimate the parameters of the Graded Response Model (which, for binary items, became the two Parameter Logistic Model). This model expresses the probability of responding in a certain way as a function of the properties of both the item and the individual’s latent trait.



Let’s start by seeing if there is only one dominant latent trait that drives the responses observed for each factor. To do that we can use the following scree plot of the eigenvalues and the subsequent table: Chart, line chart

Description automatically generated

As we can see from the images above the first eigen value is much higher than the others and it explains almost the 25% of the total variabilty, hence the model hypothesis is satisfied and so a unidimensional model is reasonable in this example.

Now we can have a look at the parameter estimations, in doing this we will focus on the threshold parameter, which represent the difficulty of the items (the higher the threshold the higher the difficulty of that item) and also on the slope, which is the discrimination parameter (the higher the slope the better the item is in differentiating subjects), actually the value of the slope represents also the amount of information provided by a certain item, hence items with an higher slope are more informative than items with a lower slope.

To preserve the reader's eyes all the easier items will be higlighted in Green, the harder will be highlited in Red, and the one with higher discrimination power will be highlited in Yellow.

| **Item Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Response Model** | **Item** | **Label** | **Parameter** | **Estimate** | **Standard Error** | **Pr > |t|** |
| **Graded** | **WINE\_PREFERENCE** | WINE\_PREFERENCE | **Threshold 1** | -1.59165 | 0.16129 | <.0001 |
|  |  |  | **Threshold 2** | -0.86753 | 0.10868 | <.0001 |
|  |  |  | **Threshold 3** | 0.03422 | 0.09345 | 0.3571 |
|  |  |  | **Slope** | 2.83221 | 0.45968 | <.0001 |
|  | **BEER\_PREFERENCE** | BEER\_PREFERENCE | **Threshold 1** | -2.65182 | 0.59002 | <.0001 |
|  |  |  | **Threshold 2** | -0.82768 | 0.27311 | 0.0012 |
|  |  |  | **Threshold 3** | 1.15645 | 0.31895 | 0.0001 |
|  |  |  | **Slope** | 0.64378 | 0.14165 | <.0001 |
|  | **SOFT\_PREFERENCE** | SOFT\_PREFERENCE | **Threshold 1** | 2.78856 | 0.66241 | <.0001 |
|  |  |  | **Threshold 2** | 0.20342 | 0.24532 | 0.2035 |
|  |  |  | **Threshold 3** | -1.95424 | 0.49841 | <.0001 |
|  |  |  | **Slope** | -0.56110 | 0.13133 | <.0001 |
|  | **COCKTAIL\_PREFERENCE** | COCKTAIL\_PREFERENCE | **Threshold 1** | 39.62510 | 142.48227 | 0.3905 |
|  |  |  | **Threshold 2** | 2.65780 | 10.14578 | 0.3967 |
|  |  |  | **Threshold 3** | -31.50778 | 113.46249 | 0.3906 |
|  |  |  | **Slope** | -0.03403 | 0.12249 | 0.3906 |
|  | **WHITE\_WINE** | WHITE\_WINE | **Threshold 1** | -3.54161 | 0.80711 | <.0001 |
|  |  |  | **Threshold 2** | -1.16692 | 0.33262 | 0.0002 |
|  |  |  | **Threshold 3** | 1.59408 | 0.40163 | <.0001 |
|  |  |  | **Slope** | 0.62935 | 0.14370 | <.0001 |
|  | **ROSE\_WINE** | ROSE\_WINE | **Threshold 1** | -3.62641 | 1.23381 | 0.0016 |
|  |  |  | **Threshold 2** | 0.78172 | 0.41878 | 0.0310 |
|  |  |  | **Threshold 3** | 4.78673 | 1.60011 | 0.0014 |
|  |  |  | **Slope** | 0.38825 | 0.12966 | 0.0014 |
|  | **RED\_WINE** | RED\_WINE | **Threshold 1** | -1.79584 | 0.23083 | <.0001 |
|  |  |  | **Threshold 2** | -0.84169 | 0.14423 | <.0001 |
|  |  |  | **Threshold 3** | 0.54945 | 0.13461 | <.0001 |
|  |  |  | **Slope** | 1.53415 | 0.22547 | <.0001 |
|  | **SPARKLING\_WINE** | SPARKLING\_WINE | **Threshold 1** | -3.29599 | 1.07951 | 0.0011 |
|  |  |  | **Threshold 2** | -0.39311 | 0.35327 | 0.1329 |
|  |  |  | **Threshold 3** | 3.63559 | 1.17117 | 0.0010 |
|  |  |  | **Slope** | 0.39789 | 0.12643 | 0.0008 |
|  | **SWEET\_WINE** | SWEET\_WINE | **Threshold 1** | -2.62255 | 1.00459 | 0.0045 |
|  |  |  | **Threshold 2** | 1.04644 | 0.50787 | 0.0197 |
|  |  |  | **Threshold 3** | 4.43376 | 1.59291 | 0.0027 |
|  |  |  | **Slope** | 0.35040 | 0.12554 | 0.0026 |
|  | **WINE\_KNOWLEDGE** | WINE\_KNOWLEDGE | **Threshold 1** | -0.67394 | 0.12467 | <.0001 |
|  |  |  | **Threshold 2** | 1.57147 | 0.18934 | <.0001 |
|  |  |  | **Threshold 3** | 1.69023 | 0.20065 | <.0001 |
|  |  |  | **Slope** | 1.70662 | 0.23243 | <.0001 |
|  | **BUYING\_EXPERIENCE** | BUYING\_EXPERIENCE | **Threshold 1** | -0.92988 | 0.22021 | <.0001 |
|  |  |  | **Threshold 2** | -2.60706 | 0.50161 | <.0001 |
|  |  |  | **Threshold 3** | -3.30831 | 0.66036 | <.0001 |
|  |  |  | **Threshold 4** | -4.62985 | 0.99061 | <.0001 |
|  |  |  | **Threshold 5** | -4.82015 | 1.04294 | <.0001 |
|  |  |  | **Slope** | -0.82683 | 0.17443 | <.0001 |
|  | **WINE\_BOTTLES** | WINE\_BOTTLES | **Threshold 1** | -1.18017 | 0.17053 | <.0001 |
|  |  |  | **Threshold 2** | -0.79653 | 0.14456 | <.0001 |
|  |  |  | **Threshold 3** | 0.95395 | 0.15407 | <.0001 |
|  |  |  | **Threshold 4** | 1.82755 | 0.22763 | <.0001 |
|  |  |  | **Threshold 5** | 2.35593 | 0.29193 | <.0001 |
|  |  |  | **Threshold 6** | 3.00070 | 0.39811 | <.0001 |
|  |  |  | **Slope** | 1.43101 | 0.19037 | <.0001 |
|  | **SUPERMARKET** | SUPERMARKET | **Threshold 1** | 6.01085 | 2.00352 | 0.0013 |
|  |  |  | **Threshold 2** | 2.78711 | 0.98738 | 0.0024 |
|  |  |  | **Threshold 3** | -1.72351 | 0.61566 | 0.0026 |
|  |  |  | **Slope** | -0.40504 | 0.13528 | 0.0014 |
|  | **WINE\_SHOP** | WINE\_SHOP | **Threshold 1** | -1.36555 | 0.27511 | <.0001 |
|  |  |  | **Threshold 2** | 1.49949 | 0.30453 | <.0001 |
|  |  |  | **Threshold 3** | 3.30606 | 0.57655 | <.0001 |
|  |  |  | **Slope** | 0.80991 | 0.14654 | <.0001 |
|  | **ONLINE\_SHOP** | ONLINE\_SHOP | **Threshold 1** | 1.71124 | 0.24675 | <.0001 |
|  |  |  | **Threshold 2** | 2.64462 | 0.38779 | <.0001 |
|  |  |  | **Threshold 3** | 3.92050 | 0.67427 | <.0001 |
|  |  |  | **Slope** | 1.37631 | 0.25431 | <.0001 |
|  | **GRAPE\_ORIGIN** | GRAPE\_ORIGIN | **Threshold 1** | -3.54765 | 0.69959 | <.0001 |
|  |  |  | **Threshold 2** | -1.68537 | 0.34120 | <.0001 |
|  |  |  | **Threshold 3** | 1.40101 | 0.31619 | <.0001 |
|  |  |  | **Slope** | 0.75927 | 0.14747 | <.0001 |
|  | **GRAPE\_VARIETY** | GRAPE\_VARIETY | **Threshold 1** | -2.89396 | 0.55587 | <.0001 |
|  |  |  | **Threshold 2** | -1.31580 | 0.27748 | <.0001 |
|  |  |  | **Threshold 3** | 1.74595 | 0.35360 | <.0001 |
|  |  |  | **Slope** | 0.79652 | 0.15153 | <.0001 |
|  | **BUDGET\_FRIENDLY** | BUDGET\_FRIENDLY | **Threshold 1** | 209.77422 | 1669 | 0.4500 |
|  |  |  | **Threshold 2** | 94.36399 | 751.44903 | 0.4500 |
|  |  |  | **Threshold 3** | -54.91275 | 436.51871 | 0.4499 |
|  |  |  | **Slope** | -0.01619 | 0.12885 | 0.4500 |
|  | **BRAND\_AWARNESS** | BRAND\_AWARNESS | **Threshold 1** | 8.29808 | 3.74706 | 0.0134 |
|  |  |  | **Threshold 2** | 3.78289 | 1.75350 | 0.0155 |
|  |  |  | **Threshold 3** | -5.32198 | 2.38460 | 0.0128 |
|  |  |  | **Slope** | -0.29052 | 0.13132 | 0.0135 |
|  | **VINTAGE** | VINTAGE | **Threshold 1** | -3.01494 | 0.73493 | <.0001 |
|  |  |  | **Threshold 2** | 1.54781 | 0.44167 | 0.0002 |
|  |  |  | **Threshold 3** | 3.47411 | 0.84594 | <.0001 |
|  |  |  | **Slope** | 0.54007 | 0.13171 | <.0001 |
|  | **LABEL\_INFO** | LABEL\_INFO | **Threshold 1** | 13.67609 | 11.88621 | 0.1250 |
|  |  |  | **Threshold 2** | 4.21622 | 3.81859 | 0.1348 |
|  |  |  | **Threshold 3** | -10.12689 | 8.71877 | 0.1227 |
|  |  |  | **Slope** | -0.14228 | 0.12337 | 0.1244 |
|  | **PACKAGING** | PACKAGING | **Threshold 1** | 19.57622 | 30.90542 | 0.2632 |
|  |  |  | **Threshold 2** | -10.11127 | 15.86080 | 0.2619 |
|  |  |  | **Threshold 3** | -27.62187 | 43.44567 | 0.2625 |
|  |  |  | **Slope** | -0.08080 | 0.12707 | 0.2624 |
|  | **PROMOTION** | PROMOTION | **Threshold 1** | 7.78455 | 5.64390 | 0.0839 |
|  |  |  | **Threshold 2** | -4.49819 | 3.23816 | 0.0824 |
|  |  |  | **Threshold 3** | -9.77802 | 7.00982 | 0.0815 |
|  |  |  | **Slope** | -0.17566 | 0.12590 | 0.0815 |
|  | **BOTTLE\_BUDGET** | BOTTLE\_BUDGET | **Threshold 1** | -8.15993 | 10.05835 | 0.2086 |
|  |  |  | **Threshold 2** | -16.25513 | 20.06395 | 0.2089 |
|  |  |  | **Threshold 3** | -25.74326 | 31.88792 | 0.2097 |
|  |  |  | **Slope** | -0.13231 | 0.16395 | 0.2098 |
|  | **BUYING\_FREQUENCY** | BUYING\_FREQUENCY | **Threshold 1** | -11.77202 | 8.35868 | 0.0795 |
|  |  |  | **Threshold 2** | -9.06877 | 6.42516 | 0.0791 |
|  |  |  | **Threshold 3** | 6.22357 | 4.46522 | 0.0817 |
|  |  |  | **Threshold 4** | 15.42351 | 10.99089 | 0.0803 |
|  |  |  | **Slope** | 0.18333 | 0.13052 | 0.0801 |
|  | **ETNA\_BUYING** | ETNA\_BUYING | **Threshold 1** | -2.00352 | 0.26918 | <.0001 |
|  |  |  | **Threshold 2** | 0.26302 | 0.13089 | 0.0222 |
|  |  |  | **Slope** | 1.43941 | 0.22011 | <.0001 |
|  | **ETNA\_PREFERENCE** | ETNA\_PREFERENCE | **Threshold 1** | 1166 | 66158 | 0.4930 |
|  |  |  | **Threshold 2** | 432.39270 | 24529 | 0.4930 |
|  |  |  | **Threshold 3** | -1099 | 62369 | 0.4930 |
|  |  |  | **Slope** | -0.00263 | 0.14916 | 0.4930 |
|  | **ETNA\_FLAVOR** | ETNA\_FLAVOR | **Threshold 1** | -8.46354 | 3.30396 | 0.0052 |
|  |  |  | **Threshold 2** | -4.25854 | 1.61658 | 0.0042 |
|  |  |  | **Threshold 3** | 4.39410 | 1.69737 | 0.0048 |
|  |  |  | **Slope** | 0.40927 | 0.15898 | 0.0050 |
|  | **SICILIAN\_EXCELLENCES** | SICILIAN\_EXCELLENCES | **Threshold 1** | -5.91337 | 1.79720 | 0.0005 |
|  |  |  | **Threshold 2** | -4.18552 | 1.24789 | 0.0004 |
|  |  |  | **Threshold 3** | 2.96824 | 0.90152 | 0.0005 |
|  |  |  | **Slope** | 0.50614 | 0.15423 | 0.0005 |
|  | **ETNA\_EXPENSIVE** | ETNA\_EXPENSIVE | **Threshold 1** | -4.49612 | 1.31150 | 0.0003 |
|  |  |  | **Threshold 2** | 3.33060 | 0.97242 | 0.0003 |
|  |  |  | **Threshold 3** | 6.90742 | 2.04994 | 0.0004 |
|  |  |  | **Slope** | 0.53546 | 0.16098 | 0.0004 |
|  | **ETNA\_QUALITY** | ETNA\_QUALITY | **Threshold 1** | -3.63592 | 0.63831 | <.0001 |
|  |  |  | **Threshold 2** | -1.85019 | 0.29658 | <.0001 |
|  |  |  | **Threshold 3** | 1.65212 | 0.27171 | <.0001 |
|  |  |  | **Slope** | 1.08537 | 0.18407 | <.0001 |
|  | **ETNA\_RECOMMENDATION** | ETNA\_RECOMMENDATION | **Threshold 1** | 66.97502 | 184.43214 | 0.3582 |
|  |  |  | **Threshold 2** | 47.31564 | 130.15830 | 0.3581 |
|  |  |  | **Threshold 3** | 18.91056 | 51.96661 | 0.3580 |
|  |  |  | **Slope** | -0.05795 | 0.15959 | 0.3583 |
| **TwoP** | **WINE\_TASTING** | WINE\_TASTING | **Difficulty** | 0.17557 | 0.10649 | 0.0496 |
|  |  |  | **Slope** | 2.11417 | 0.34795 | <.0001 |
|  | **WINERY\_VISIT** | WINERY\_VISIT | **Difficulty** | -0.43781 | 0.12847 | 0.0003 |
|  |  |  | **Slope** | 1.43703 | 0.26620 | <.0001 |
|  | **WINE\_COURSE** | WINE\_COURSE | **Difficulty** | 1.21970 | 0.15715 | <.0001 |
|  |  |  | **Slope** | 2.06060 | 0.36829 | <.0001 |
|  | **PARTY** |  | **Difficulty** | -10.33400 | 12.87826 | 0.2111 |
|  |  |  | **Slope** | -0.14595 | 0.18279 | 0.2123 |
|  | **GIFT** |  | **Difficulty** | 1.69885 | 0.42342 | <.0001 |
|  |  |  | **Slope** | 0.69337 | 0.17282 | <.0001 |
|  | **HOME** |  | **Difficulty** | -0.44917 | 0.14870 | 0.0013 |
|  |  |  | **Slope** | 1.15444 | 0.21601 | <.0001 |
|  | **TASTE** |  | **Difficulty** | 5.00236 | 2.28163 | 0.0142 |
|  |  |  | **Slope** | 0.47359 | 0.22905 | 0.0193 |
|  | **ETNA\_DOC** | ETNA\_DOC | **Difficulty** | -0.26353 | 0.13698 | 0.0272 |
|  |  |  | **Slope** | 1.21552 | 0.22333 | <.0001 |

According to the table above we can identify the easiest and the hardest items. Let’s start with the easiest ones:

* Packaging
* Bottle\_Budget
* Buying\_Frequency
* Party

Now let’s see the hardest ones:

* Cocktail\_Preference
* Budget\_Friendly
* Label\_Info
* Etna\_Preference
* Etna\_Recommendation

Now let’s analize also the slope parameter, which represents the discrimination parameter. Recalling that the slope parameter identifies the amount of information described by each item we can analize it looking at the following information curves:

A picture containing histogram

Description automatically generated

Histogram

Description automatically generated

Chart, histogram

Description automatically generated

Histogram

Description automatically generated

Graphical user interface, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated with medium confidence

As we can see from these plots, all the items identified as easy and hard provide little to no information at all . We can try to remove all these items because they are not general enough and do not provide any sort of information.

A picture containing text, different

Description automatically generatedAs we can see in the images below the item test curve do not change at all after the removal of these items.

## 4.2.2 IRT on topics

Now that we have seen the properties of the various items at a global level, we can focus on each subgroup of items, identified by each topic of our questionnaire.

In the questionnaire there are 3 main topics:

1. Wine preferences, expertise and experience
2. Wine buying experience
3. Etna DOC Wine

and 1 socio-demographic topic that we will ignore.

For each topic the same approach of the global level IRT will be used.

*4.2.2.1 Wine preferences, expertise and experience*

From the following scree plot we can identify that there is only one latent trait guiding the responses

Chart, line chart

Description automatically generated

Let’s have a look at the estimated parameters

| **Item Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Response Model** | **Item** | **Label** | **Parameter** | **Estimate** | **Standard Error** | **Pr > |t|** |
| **Graded** | **WINE\_PREFERENCE** | WINE\_PREFERENCE | **Threshold 1** | -1.50560 | 0.13448 | <.0001 |
|  |  |  | **Threshold 2** | -0.70713 | 0.09534 | <.0001 |
|  |  |  | **Threshold 3** | 0.11362 | 0.08044 | 0.0789 |
|  |  |  | **Slope** | 5.69107 | 1.66047 | 0.0003 |
|  | **BEER\_PREFERENCE** | BEER\_PREFERENCE | **Threshold 1** | -1.91935 | 0.32388 | <.0001 |
|  |  |  | **Threshold 2** | -0.60688 | 0.17825 | 0.0003 |
|  |  |  | **Threshold 3** | 0.82403 | 0.19782 | <.0001 |
|  |  |  | **Slope** | 0.96219 | 0.15966 | <.0001 |
|  | **SOFT\_PREFERENCE** | SOFT\_PREFERENCE | **Threshold 1** | 3.50717 | 1.06623 | 0.0005 |
|  |  |  | **Threshold 2** | 0.24857 | 0.31266 | 0.2133 |
|  |  |  | **Threshold 3** | -2.47245 | 0.78526 | 0.0008 |
|  |  |  | **Slope** | -0.43257 | 0.13042 | 0.0005 |
|  | **COCKTAIL\_PREFERENCE** | COCKTAIL\_PREFERENCE | **Threshold 1** | -6.87997 | 4.51098 | 0.0636 |
|  |  |  | **Threshold 2** | -0.42631 | 0.71702 | 0.2761 |
|  |  |  | **Threshold 3** | 5.51072 | 3.61579 | 0.0637 |
|  |  |  | **Slope** | 0.19615 | 0.12740 | 0.0618 |
|  | **WHITE\_WINE** | WHITE\_WINE | **Threshold 1** | -2.35695 | 0.35912 | <.0001 |
|  |  |  | **Threshold 2** | -0.76947 | 0.17923 | <.0001 |
|  |  |  | **Threshold 3** | 1.06039 | 0.20534 | <.0001 |
|  |  |  | **Slope** | 1.05185 | 0.16464 | <.0001 |
|  | **ROSE\_WINE** | ROSE\_WINE | **Threshold 1** | -2.29161 | 0.50341 | <.0001 |
|  |  |  | **Threshold 2** | 0.50536 | 0.23352 | 0.0152 |
|  |  |  | **Threshold 3** | 3.00009 | 0.64319 | <.0001 |
|  |  |  | **Slope** | 0.64874 | 0.13767 | <.0001 |
|  | **RED\_WINE** | RED\_WINE | **Threshold 1** | -1.48574 | 0.15549 | <.0001 |
|  |  |  | **Threshold 2** | -0.64289 | 0.10826 | <.0001 |
|  |  |  | **Threshold 3** | 0.46628 | 0.10121 | <.0001 |
|  |  |  | **Slope** | 2.51156 | 0.32044 | <.0001 |
|  | **SPARKLING\_WINE** | SPARKLING\_WINE | **Threshold 1** | -2.48933 | 0.63269 | <.0001 |
|  |  |  | **Threshold 2** | -0.30987 | 0.25859 | 0.1154 |
|  |  |  | **Threshold 3** | 2.71771 | 0.68172 | <.0001 |
|  |  |  | **Slope** | 0.54458 | 0.13307 | <.0001 |
|  | **SWEET\_WINE** | SWEET\_WINE | **Threshold 1** | -1.71241 | 0.47612 | 0.0002 |
|  |  |  | **Threshold 2** | 0.70333 | 0.28704 | 0.0071 |
|  |  |  | **Threshold 3** | 2.90880 | 0.72076 | <.0001 |
|  |  |  | **Slope** | 0.55342 | 0.13584 | <.0001 |
|  | **WINE\_KNOWLEDGE** | WINE\_KNOWLEDGE | **Threshold 1** | -0.75610 | 0.16687 | <.0001 |
|  |  |  | **Threshold 2** | 1.87656 | 0.27827 | <.0001 |
|  |  |  | **Threshold 3** | 2.01532 | 0.29694 | <.0001 |
|  |  |  | **Slope** | 1.21819 | 0.19507 | <.0001 |
| **TwoP** | **WINE\_TASTING** | WINE\_TASTING | **Difficulty** | 0.29448 | 0.13185 | 0.0128 |
|  |  |  | **Slope** | 1.33426 | 0.24456 | <.0001 |
|  | **WINERY\_VISIT** | WINERY\_VISIT | **Difficulty** | -0.61752 | 0.22668 | 0.0032 |
|  |  |  | **Slope** | 0.75703 | 0.17666 | <.0001 |
|  | **WINE\_COURSE** | WINE\_COURSE | **Difficulty** | 1.51179 | 0.26747 | <.0001 |
|  |  |  | **Slope** | 1.30343 | 0.29295 | <.0001 |

This section doesn’t seem to have neither too easy itmes nor to difficult ones, the only item that have more “extreme” threshold values is Cocktail\_Preference, but also here the values are relatively balanced.

Let’s see how the discrimination values behave

Graphical user interface, histogram

Description automatically generated with medium confidence

Histogram

Description automatically generated

A picture containing chart

Description automatically generated

As we can see here half of the items provide a small contribution to the analysis and 3 of them apport litteraly 0 contribution. Let’s see if we can remove those 3 items without losing information

Histogram

Description automatically generated with medium confidence

As we can clearly see, the shape remains almost the same, hence it can make sense to remove these items from this topic.

*4.2.2.2 Wine buying experience*

The items that fall into this topic are:

* BUYING\_EXPERIENCE
* WINE\_BOTTLES
* SUPERMARKET
* WINE\_SHOP
* ONLINE\_SHOP
* GRAPE\_ORIGIN
* GRAPE\_VARIETY
* BUDGET\_FRIENDLY
* BRAND\_AWARNESS
* VINTAGE
* LBEL\_INFO
* PACKAGING
* PROMOTION
* BOTTLE\_BUDGET
* BUYING\_FREQUENCY
* PARTY
* GIFT
* HOME
* TASTE

Analize the scree plot for this topic we find out that there are more than one driving force that influece the responses of the subjects (we can identify that considering that the first two eigenvalues are really close).

Chart, line chart

Description automatically generated

Let’s check how the items relate to each of the 2 main factors.

| **Rotated Slope Matrix** | | |
| --- | --- | --- |
|  | **\_Factor1** | **\_Factor2** |
| **BUYING\_EXPERIENCE** | -0.28159 | 0.76757 |
| **WINE\_BOTTLES** | 0.31072 | 0.52798 |
| **SUPERMARKET** | -1.55931 | 0.14830 |
| **WINE\_SHOP** | 2.14141 | -0.23166 |
| **ONLINE\_SHOP** | 0.89898 | 0.01597 |
| **GRAPE\_ORIGIN** | 1.22596 | -0.78631 |
| **GRAPE\_VARIETY** | 1.42912 | -0.47651 |
| **BUDGET\_FRIENDLY** | -0.96266 | -0.48930 |
| **BRAND\_AWARNESS** | -0.86122 | -1.54697 |
| **VINTAGE** | 0.46446 | -1.09839 |
| **LABEL\_INFO** | -0.19131 | -0.97166 |
| **PACKAGING** | -0.79830 | -0.56321 |
| **PROMOTION** | -1.19528 | -0.20459 |
| **BOTTLE\_BUDGET** | -0.46467 | 1.66485 |
| **BUYING\_FREQUENCY** | -0.14678 | 0.28150 |
| **PARTY** | -0.53597 | -0.06288 |
| **GIFT** | 0.23626 | 0.43151 |
| **HOME** | 0.17032 | 0.71065 |
| **TASTE** | -0.02893 | 1.18769 |

In order to be sure of this division of the items let’s apply a confirmatory model.

| **Slope Matrix Estimate/StdErr/p-value** | | |
| --- | --- | --- |
|  | **Factor1** | **Factor2** |
| **BUYING\_EXPERIENCE** | 0.00000 | 0.51159  0.16259  0.00083 |
| **WINE\_BOTTLES** | 0.00000 | -3.15877  0.88804  0.00019 |
| **SUPERMARKET** | 1.57173  0.24741  <.00001 | 0.00000 |
| **WINE\_SHOP** | -2.47685  0.43741  <.00001 | 0.00000 |
| **ONLINE\_SHOP** | -0.75966  0.25050  0.00121 | 0.00000 |
| **GRAPE\_ORIGIN** | -1.15961  0.19908  <.00001 | 0.00000 |
| **GRAPE\_VARIETY** | -1.36410  0.21743  <.00001 | 0.00000 |
| **BUDGET\_FRIENDLY** | 0.80295  0.18155  <.00001 | 0.00000 |
| **BRAND\_AWARNESS** | 0.00000 | 0.60214  0.15489  0.00005 |
| **VINTAGE** | 0.00000 | 0.04228  0.13616  0.37808 |
| **LABEL\_INFO** | 0.00000 | 0.52122  0.13569  0.00006 |
| **PACKAGING** | 0.73185  0.16471  <.00001 | 0.00000 |
| **PROMOTION** | 1.15712  0.19748  <.00001 | 0.00000 |
| **BOTTLE\_BUDGET** | 0.00000 | -0.94409  0.24067  0.00004 |
| **BUYING\_FREQUENCY** | 0.00000 | -0.42096  0.14295  0.00162 |
| **PARTY** | 0.46113  0.19479  0.00896 | 0.00000 |
| **GIFT** | 0.00000 | -1.01689  0.24712  0.00002 |
| **HOME** | 0.00000 | -3.05037  0.84937  0.00016 |
| **TASTE** | 0.00000 | -1.07522  0.35512  0.00123 |

As we can see, with a 0.05 confidence level, all our “classifications” are statistically significant but one: Vintage.

Re-running the test but assignign Vintage to the first factor we get that it has a slope matrix estimate of -0.58537 with a standard error of 0.15893 and a P-Value of 0.00012, therefore we can assign Vintage to the first factor.

Let’s run IRT individually on each of these factors.

*4.2.2.3 Factor1*