

Consumer Perception Impacts on Olive Oil Consumption Choice: A Machine Learning Approach

Youmin Li

Department of Food and Resource Economics, University of Florida

liyoumin@ufl.edu

October 31, 2025

Background

Extra virgin olive oil (EVOO) offers documented sensory and health advantages, yet adoption remains uneven relative to refined olive oil (ROO). Standard models struggle with dozens of collinear Likert items and heterogeneous preferences. This project builds an integrated machine learning (ML) + partial least squares structural equation model (PLS-SEM) pipeline to (i) recover low-noise perception constructs, (ii) segment consumers, and (iii) quantify direct, indirect, and moderated pathways from perceptions to consumption choices. Insights inform labeling, targeting, and sustainable market development for EVOO.

Data Description

Table 1: Socio-demographic characteristics of the Spanish population and the study sample.

Characteristic	Sample (%)	Spain (%)
Age		
18–35 years	27	24
36–50 years	32	29
>50 years	41	47
Gender		
Male	46	49
Female	54	51
Education		
Primary education	12	36
Secondary education	41	23
University	34	41

The study employs survey data from 1,031 Spanish consumers collected via a discrete choice experiment and Likert-scale questionnaire developed by the University of Córdoba and the Spanish Ministry of Agriculture. Variables include consumption frequency of EVOO/ROO, perceived health benefits, taste preference, price gap, and trust. Demographic covariates include gender, income, education, and region. The sample closely

matches national distributions, making it representative of Spain’s consumer population.

Research Questions

- **RQ1:** Which perception constructs most strongly predict EVOO (vs. ROO) choice?
- **RQ2:** Do effects operate indirectly via trust/attitudes (mediation)? Are they stronger for specific segments (moderation)?

Methodology

The analysis proceeds in two stages. **Stage I (Unsupervised learning)** applies exploratory factor analysis (EFA-PCA) to condense correlated perception items into latent constructs, and clustering (k-means/Gaussian mixture) to detect respondent segments and outliers. **Stage II (Supervised learning)** links these constructs to consumption outcomes using penalized logistic regression (LASSO) and random forest classifiers, validated via ten-fold cross-validation. Partial Least Squares Structural Equation Modeling (PLS-SEM) is further employed to explore mediating and moderating effects among latent factors, enabling estimation of both direct and indirect influences on consumption choice.

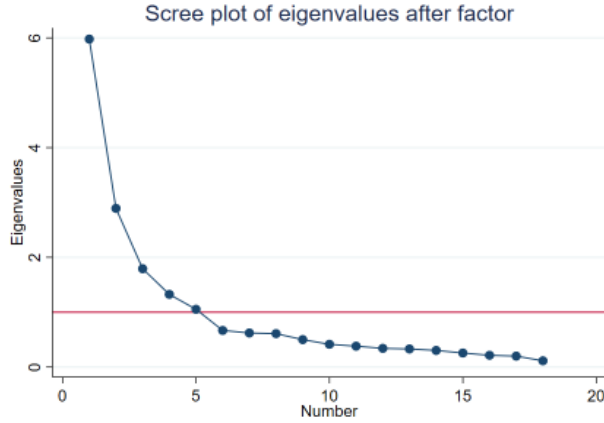


Figure 1: Screen plot

Preliminary Results

Factor analysis yielded five latent components ($KMO = 0.88$, Bartlett’s $p < 0.001$) capturing *trust*, *taste*, *price perception*, *health belief*, and *consumption behavior*. Reliability metrics indicate strong internal consistency (e.g., $\alpha_{EVOOTrust} = 0.83$, $\alpha_{ROOTrust} = 0.90$). Early PLS-SEM and ML models reveal that perceived quality and certification trust, followed by taste imagery, are the most significant predictors of EVOO consumption.

Cross-validated LASSO and random forest models achieve AUC values between 0.78–0.83, suggesting robust predictive power. Heterogeneity analysis shows that high-income and frequent users are more influenced by quality perceptions, while lower-income consumers emphasize price sensitivity.

Additional Information

Table 2: Description of variables used in the olive oil consumption study.

Variable	Description
EVOO Use	Weekly uses of extra virgin olive oil (EVOO) in the household, measured as number of meal uses per week.
ROO Use	Weekly uses of refined olive oil (ROO) in the household, measured as number of meal uses per week.
EVOO Con	Monthly consumption of EVOO in the household, in liters per month per person.
ROO Con	Monthly consumption of ROO in the household, in liters per month per person.
Trust ROO1	Your degree of trust in ROO, 1–7 Likert scale.
Trust ROO2	The degree to which you need ROO, 1–7 Likert scale.
Trust ROO3	The degree to which you recommend ROO, 1–7 Likert scale.
Trust ROO4	The perceived quality of ROO, 1–7 Likert scale.
Trust EVOO1	Your degree of trust in EVOO, 1–7 Likert scale.
Trust EVOO2	The degree to which you need EVOO, 1–7 Likert scale.
Trust EVOO3	The degree to which you recommend EVOO, 1–7 Likert scale.
Trust EVOO4	The perceived quality of EVOO, 1–7 Likert scale.
Taste 1	“EVOO is less versatile in the kitchen due to its taste” (reverse-coded), 1–7 Likert scale.
Taste 2	“I prefer a mild and light-flavored olive oil” (reverse-coded), 1–7 Likert scale.
Taste 3	“EVOO adds too strong a flavor for most dishes” (reverse-coded), 1–7 Likert scale.
Price 1	“Extra virgin olive oil has a suitable price,” 1–7 Likert scale.
Price 2	“The price gap between EVOO and ROO is small,” 1–7 Likert scale.
Price 3	“Considering annual food outlay, EVOO is a cheap product,” 1–7 Likert scale.
Price 4	“Considering its features, ROO has a high price,” 1–7 Likert scale.
Health 1	“EVOO and ROO have the same health benefits,” 1–7 Likert scale.
Health 2	“EVOO is not as good for frying as ROO,” 1–7 Likert scale.
Health 3	“EVOO and ROO have the same features except for taste,” 1–7 Likert scale.