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

Youmin Li

University of Florida
Department of Food and Resource Economics

December 2025

Background

- Olive oil markets are increasingly get attention due to healthy attribution. Consumers' choices between extra virgin olive oil (EVOO) and refined olive oil (ROO) are influenced by these multifaceted perceptions.
- **Research gaps:** Traditional studies focus on a few variables, but modern surveys collect dozens of perception indicators. There is a need to distill complex survey data to understand how consumer perceptions drive EVOO vs. ROO preferences.

Motivation

- **Leverage machine learning amis**
  - Reduce dimensionality and denoise multi-item perception constructs.
  - Identify latent consumer segments; and predict olive oil consumption outcomes.
  - Reveal how perceptions of taste, price, health, and trust impact the choice of EVOO, and how these effects vary across consumer segments.
- **Contribution**
  - Provide a novel machine learning framework to study consumer choices.
  - Evaluated predication accuracy of different models (LASSO, Logistic, Random forest, GAM), and reveal perception factors' importance to olive oil consumption choice.

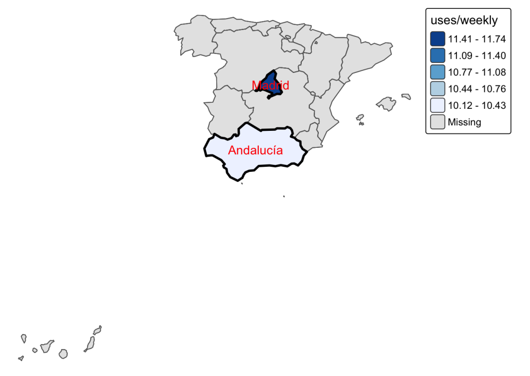

Figure 1: Average uses of EVOO per week per household

Data overview

- **Sample**: Cross-sectional survey of 1031 adult consumers in Spain. The sample is broadly representative of the national population, providing practical relevance to the results.
- **Data collection**: Respondents completed a choice experiment and Likert-scale questionnaire on olive oil usage and perceptions. The instrument was developed via pilot factor analysis to ensure content validity.
- **Data variables**: perception constructs (trust, price, taste, healthy perception). Demographics (Age, gender, income level, education, and region).

EVOO Use	Weekly uses of EVOO in the household, number of meal uses per week.
ROO Use	Weekly uses of ROO in the household, number of meal uses per week.
EVOO Con	Monthly consumption of EVOO in the household, liters/month/person.
ROO Con	Monthly consumption of ROO in the household, liters/month/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), 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.
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.

Methods: Unsupervised ML

- EFA-PCA: reduce numerous perception survey items into a smaller set of latent factors. Addressed collinearity and grouped related perceptions.
- Cluster: Used K-means clustering on respondents' factor scores to discover latent consumer segments. Number of clusters was chosen via silhouette analysis.
- Path Modeling: Built a Partial Least Squares Structural Equation Model (PLS-SEM) to assesses mediating/moderating effects (e.g., how taste influences trust, which in turn influences EVOO choice).

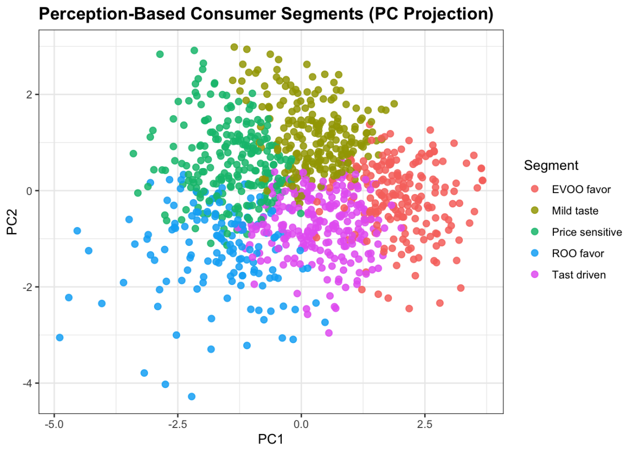

Figure 2:Consumers segments

Methods: Supervised ML

Predictive modeling

- Implemented supervised learning to predict EVOO vs. ROO preference. Evaluated multiple classifiers – Logistic Regression, LASSO-penalized Logistic, Random Forest, Gradient Boosted Trees (XGBoost), and a GAM – using stratified 10-fold cross-validation. The LASSO model was emphasized for its balance of accuracy and interpretability.

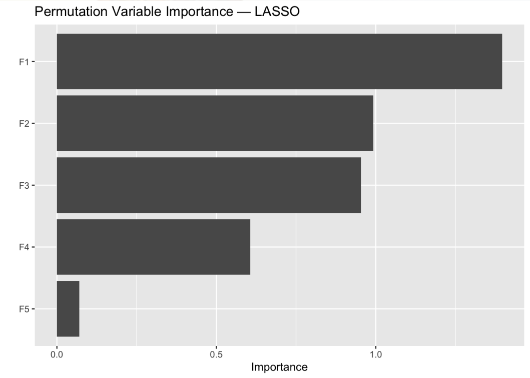

Figure 3:Factors importance

Results: PLS-SEM

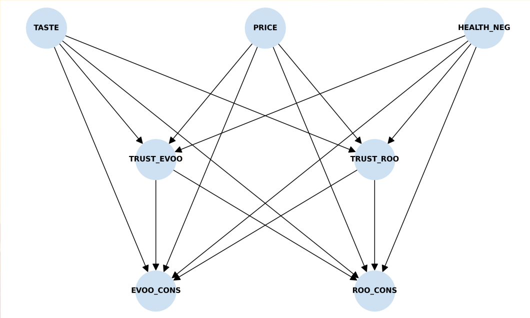

Figure 4:Structural equation model - pathway
(Note: conceptual path diagram linking perception constructs to consumption outcomes.)

- A stronger preference for taste tends to increase trust in EVOO but decrease trust in ROO (as indicated by path coefficients, e.g. Taste → Trust_EVOO was negative, Taste → Trust_ROO positive in the model). Trust in EVOO strongly increases EVOO usage, while trust in ROO increases ROO usage (and inversely affects EVOO choice). Price perception had indirect effects, and health perception showed little direct effect on choice in the presence of other factors.

Predictive Modeling Performance

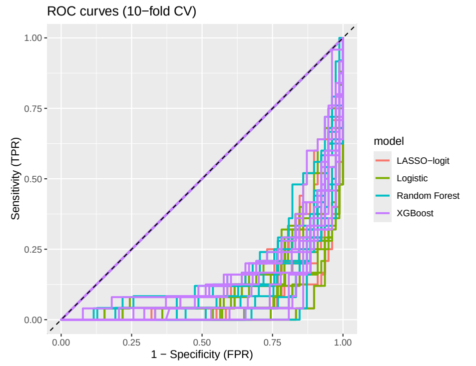

Figure 5: 10-folder cross-validation - ROC

Table: Out-of-sample performance (10-fold CV)

Model	ROC_AUC	Accuracy	Log-loss	Notes
Logistic (baseline)	0.90	0.86	0.33	–
LASSO-logit	**0.91**	**0.87**	**0.33**	λ_{min}/1SE
Random Forest	0.88	0.85	0.35	500 trees
Boosting	0.88	0.84	0.35	tuned
GAM-logit	0.89	0.87	0.32	EDF summary

Results:

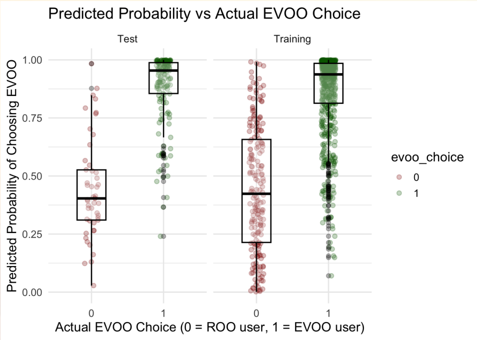

Figure 6: LASSO-logit prediction probability

Table: LASSO-logit coefficients and odds ratios for prediction

Factor	coefficients	odds ratios
F1: trust_ROO	-1.13	0.29
F2: trust$_E$*VOO*	0.92	2.22
F3: Price sensitive	-0.86	0.40
F4: Taste perception	0.60	1.73
F5: Healthy	0	1.00
Income	0.09	1.1
Age	0.09	1.09
Eduction	0.04	1.04
Region	0.17	1.18
Gender	0	1.00

Discussion and policy implications

- **Main drivers**: Perceived trust and taste imagery emerge as the key determinants of choosing EVOO. Consumers who trust EVOO's quality and are drawn by its flavor are far more likely to prefer it. Price concerns and health beliefs play secondary roles.
- **Implementation**: Targeted messaging tailored to segment-specific salience (health vs. price vs. sensory); and product positioning that aligns sensory expectations with verified quality claims are necessary.
- **Takeaways**: Future work could integrate consumer segments and demographic characters, conduct multiple group analysis (MGA) in casual empirical studies to reveal the differences among different groups, and examine substitution patterns among edible oils.

Thank You for Your Attention

Questions or Comments?

Youmin Li
Department of Food and Resource Economics
University of Florida

`liyoumin@ufl.edu`

Data and Code: github.com/liyoumin/Machine-learning-project
Presentation record: liyoumin.github.io/personalweb/teaching/js/

Thank you for your time and feedback!