Matilde Bini, Pietro Amenta, Antonello D'Ambra, Ida Camminatiello
Editors

Società Italiana di Statistica

Statistical Methods for Service Quality Evaluation

Book of short papers of IES 2019, Rome, Italy, July 4-5

SIS Società Italiana di Statistica

Cuzzolin

Matilde Bini, Pietro Amenta, Antonello D'Ambra, Ida Camminatiello
Editors

Statistical Methods for Service Quality Evaluation

Book of short papers

9th International Conference **IES 2019** - Innovation & Society -

Statistical evaluation systems at 360°: techniques, technologies and new frontiers

organized by Statistics for the Evaluation and Quality in Services Group of the Italian Statistical Society and European University of Rome

Società Italiana di Statistica

Cuzzolin

Matilde Bini - European University of Rome, Italy
Pietro Amenta - University of Sannio, Italy
Antonello D'Ambra - University of Campania "L. Vanvitelli", Italy
Ida Camminatiello - University of Campania "L. Vanvitelli", Italy
Editors

Prima Edizione: Luglio 2019

Traversa Michele Pietravalle, 8 - 80131 Napoli
Tel. 081 5451143 - Fax 081 7707340
cuzzolineditor@cuzzolin.it
www.cuzzolineditore.com

ISBN: 978-88-86638-65-4

CUZZOLIN EDITORE

A quantile regression perspective on consumer heterogeneity

Cristina Davino, Rosaria Romano, Domenico Vistocco

Abstract The main objective of the consumer analysis is to analyze the heterogeneity of preferences with respect to a predefined set of products. In some cases, consumer preferences are also related to some specific drivers in order to obtain preference models to be used in planning marketing strategies. The aim of this work is to present a strategy that allows to estimate preference models taking into account the individual differences of consumers in the liking pattern. The proposed strategy consists in using quantile regression to obtain preference models for homogeneous groups of consumers with respect to the quantile that best represents them. The strategy will be tested on data deriving from a case study on consumer's preferences for muscadine grape juices.
Abstract *L'obiettivo principale dell'analisi delle preferenze dei consumatori è analizzare l'eterogeneità delle preferenze rispetto ad un insieme predefinito di prodotti. In alcuni casi, le preferenze sono anche messe in relazione ad alcuni driver specifici, al fine di ottenere modelli di preferenza da utilizzare nella pianificazione delle strategie di marketing. Lo scopo di questo lavoro è presentare una strategia di analisi che consenta di stimare modelli di preferenza che tengano conto anche delle differenze individuali dei consumatori. La strategia proposta consiste nell'utilizzare la regressione quantile per ottenere modelli per gruppi omogenei di consumatori rispetto al quantile che meglio li rappresenta. La proposta sarà testata su dati derivanti da un caso di studio sulle preferenze dei consumatori per i succhi di uva muscadine.*

Key words: individual differences, liking model, quantile regression, segmentation

Cristina Davino, Rosaria Romano
Department of Economics and Statistics, University of Naples Federico II, e-mail: cristina.davino@unina.it, rosaria.romano@unina.it

Domenico Vistocco
Department of Political Sciences, University of Naples Federico II e-mail: domenico.vistocco@unina.it

1 Introduction

In consumer analysis it is very important to analyze consumer heterogeneity [5]. This is the main objective behind the *preference mapping*, which allows to visualize the structure of consumer preferences for a specific set of products [4]. Often the analysis of individual differences is not enough, and one also prefer to analyze how preferences depend on some specific liking drivers [2]. In this context it would be useful to obtain preference models that assess at the same time the dependence of liking on drivers and individual differences between consumers. At this aim the present paper proposes an analysis strategy based on the use of quantile regression (QR), a well-known statistical method able to go beyond the analysis of the average effects of a set of regressors on a dependent variable. QR, as introduced by Koenker and Basset [1] may be considered an extension of OLS regression because it is based on the estimation of a set of conditional quantiles of a response variable as a function of a set of regressors [3]. QR coefficients allow to analyze the contribution of each regressor on the entire conditional distribution of the dependent variable. The parameter estimates in QR linear models have the same interpretation as those of any other linear model: each coefficient represents the rate of change of the considered conditional quantile of the dependent variable per unit change in the value of a regressor, holding the others constant. Although different functional forms can be used, the paper will refer to linear regression models. Because of lack of space, the interested reader may refer to the reference literature for methodological details.
The data set used to show the proposed strategy comes from a consumer testing on 10 muscadine grape juices, in which 60 consumers expressed their preference for each product on a 9-point hedonic scale (9 = like extremely and 1 = dislike extremely) [4]. Furthermore, consumers themselves have provided a judgment on some derivers (appearance, aroma, color, flavor) using the same 9-point hedonic scale.

2 Methods and main results

The main idea of the proposal is to explore if the effect of the drivers of the overall liking differs for groups of consumers. The analysis strategy consists of three main steps. The first objective is to identify the best model for each consumer, based on the quantile that best represents each consumer. Subsequently, consumers segments are identified according to similarities in the dependence structure. Finally, a different model is estimated for each group of consumers identified in the previous step (a group can also consist of a single consumer). In the present Section, the proposed approach is described by showing at the same time the methodology and the results of the application on the grape data.

In the first step, a representative quantile is identified for each consumer as the average rank percentile of the overall liking he has expressed on the considered set of products. Let us denote with m the number of consumers (60 in the grape dataset)

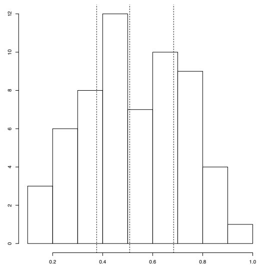

Fig. 1 Histogram of the rank percentile

and with n_g ($g = 1,\dots,m$) the number of products evaluated by each consumer (always equal to 10 in the grape dataset). The quantile representative of each group will be obtained as: $\theta_g^{best} = mean(rank\_perc(y_i))$ where $rank\_perc$ represents the location of scores in a distribution, $i = 1,\dots,n_g$ and $g = 1,\dots,m$.

The histogram in Figure 1 shows the distribution of the rank percentiles on the sample of consumers. It is worth noting that there exist a variability on the θ^{best} and thus an heterogeneity in the liking.

QR is then carried out on the whole sample using the representative quantiles, that is, the m quantiles θ_g^{best} assigned to the m consumers. Each model provides a set of coefficients, one for each driver, that can be arranged in a matrix '*consumers* $\times$ *coefficients*'. The aim of the second step is to identify consumers' segments with a different dependence structure. For this purpose, a Cluster Analysis is performed on the matrix '*consumers* $\times$ *coefficients*'. On the grape dataset, the best partition identifies seven groups of consumers characterised by a different position in the ranking of the overall liking. Table 1 describes each cluster through the following information: the minimum and the maximum value of the θ_g^{best} (θ_{min} and θ_{max}), a quantile representative of each group (the average of the θ^{best} assigned to the consumer in each cluster) and the average values of the original variables. In the final step, QR is carried out on the whole sample using the representative quantiles assigned to each cluster (θ^{best} mean) providing coefficients in Table 2 ('*' and '**' denotes coefficients significant respectively for $\alpha \leq 0.05$ and $0.05 < \alpha \leq 0.1$. The standard errors, used to evaluate the statistical significance of the coefficients, have been estimated using resampling methods (Parzen et al., 1994)). Combining the information in Table 1 and 2 it is possible to argue that Flavor is the most interesting driver especially to improve the overall liking of those less satisfied. If on one hand, on the highest part of the overall distribution the satisfaction about the Flavor is very high (average equal to 8.05 and 8.90 in clusters 6 and 7), on the other hand, in the first two clusters the averages are below the threshold of sufficiency. The importance of acting on

Table 1 Ranges of θ^{best} and description of clusters by variables and θ^{best} average

Clusters	θ min	θ max	θ^{best} mean	overall	appearance	aroma	color	flavor
1	0.10	0.21	0.15	4.30	5.45	4.62	5.30	4.05
2	0.23	0.51	0.38	6.04	6.38	6.11	6.12	5.91
3	0.54	0.62	0.58	7.19	7.14	6.80	6.96	7.10
4	0.64	0.69	0.66	7.53	7.80	7.24	7.54	7.33
5	0.71	0.73	0.72	7.79	7.87	7.56	7.81	7.40
6	0.75	0.88	0.81	8.17	8.20	8.18	8.15	8.05
7	0.98	0.98	0.98	8.90	8.00	8.50	7.80	8.90

this driver is further confirmed by a significative impact of the Flavor coefficients on the overall satisfaction, with a decreasing trend moving from the lowest to the highest quantiles. Aroma and Appearance play a significant role on the highest part of the distribution albeit at a lower intensity than the Flavor. Finally, Color is well evaluated but it has no impact on the overall satisfaction. Further developments will

Table 2 QR coefficients for the seven representative quantiles

Coefficient	$\theta = 0.15$	$\theta = 0.38$	$\theta = 0.58$	$\theta = 0.66$	$\theta = 0.72$	$\theta = 0.81$	$\theta = 0.98$
Intercept	-1.07*	0.00	0.41	1.09**	1.78**	2.54**	4.33**
Appearance	0.20	0.00	0.14	0.17 *	0.22**	0.11*	0.11
Aroma	0.07	0.00	0.04**	0.04	0.05**	0.11*	0.09*
Color	0.00	0.00	0.00	-0.03	-0.05	0.00	0.04
Flavor	0.80*	1.00	0.77*	0.68*	0.58*	0.51*	0.32*

regard a testing procedure to evaluate the the significance of the differences among the coefficients related to each cluster exploiting the classical inferential tools available in the QR framework.

References

1. Koenker, R. and Basset, G.W: Regression quantiles. Econometrica. **46**, 33–50 (1978)
2. Davino C., Romano R., Vistocco, D.: Modelling drivers of consumer liking handling consumer and product effects. Italian Journal of Applied Statistics. **30**, 359–372 (2018)
3. Davino, C., Furno, M., Vistocco, D.: Quantile Regression. Theory and Applications. Wiley Series in Probability and Statistics (2013)
4. Meullenet, J. F., Xiong, R., Findlay, C. J.: Multivariate and probabilistic analyses of sensory science problems (Vol. 25). John Wiley & Sons (2008)
5. Naes, T., Varela, P., Berget, I.: Individual Differences in Sensory and Consumer Science: Experimentation, Analysis and Interpretation. Woodhead Publishing (2018)
6. Parzen, M.I., Wei, L.J., Ying, Z.: A resampling method based on pivotal estimating functions. Biometrika. **81**, 431–350 (1994)