grostat 2012:
12th European Symposium on Statistical Methods for the Food Industry
Paris, France, (28), 29 February & 1-2 March 2012

CUB models for sensory analysis in food industry

Marica Manisera
Department of Quantitative Methods,
University of Brescia, Italy

Domenico Piccolo
Department of Statistical Science,
University of Napoli Federico II, Italy

Paola Zuccolotto
Department of Quantitative Methods,
University of Brescia, Italy

Agenda

- Introduction
- CUB Models *in brief*
- Case Study
- Conclusions

Introduction

Consumers and experts' **preferences** on the sensory attributes of products are important to producers in food industry

We need appropriate **sensory analyses** and correct **statistical methods**

This finally has a **great impact** upon food quality and industrial competitiveness

Sensory data

To collect **sensory data**, subjects are asked to rate or rank different products on the basis of some sensory descriptors by expressing their perceptions on hedonic response scales

We obtain **ordinal data**

Sensory data

Ordinal data modelling

- Regression models for ordinal data

  (McCullagh, 1980)

- Generalized linear models

  (McCullagh and Nelder, 1989)

- **CUB models**

  (Piccolo, 2003)

CUB models

The response of each subject is interpreted as the combination of

- a **feeling** attitude towards the food
- an intrinsic **uncertainty** related to the circumstances surrounding the discrete choice

CUB models

The feeling and uncertainty components are considered in the CUB models by a mixture of two random variables

$$Pr\left(R=r\right)=\pi\left[\binom{m-1}{r-1}\xi^{m-r}\left(1-\xi\right)^{r-1}\right]+\left(1-\pi\right)\left[\frac{1}{m}\right],\quad r=1,2,\ldots,m$$

feeling **uncertainty**

where $\pi \in (0,1]$ and $\xi \in [0,1]$.

Thus the parametric space is the (left open) unit square

CUB models

Parameters are related to the latent components of the responses:

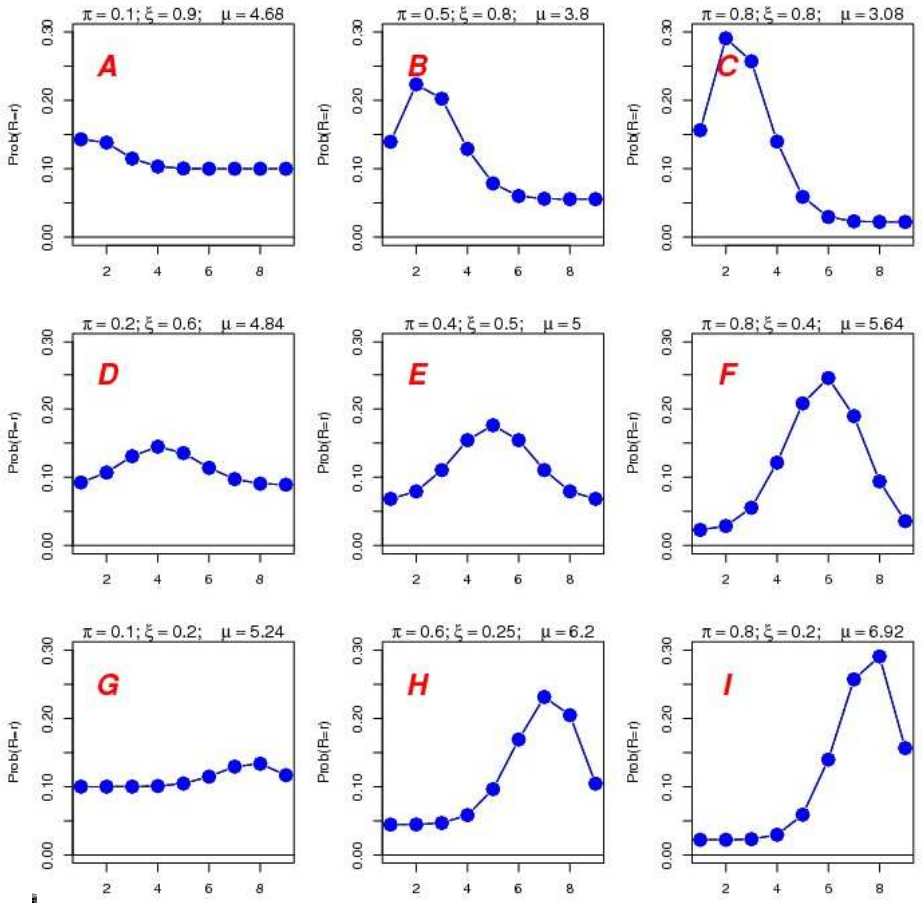
$1 - \xi \rightarrow$ sensory satisfaction (**feeling**) with the product

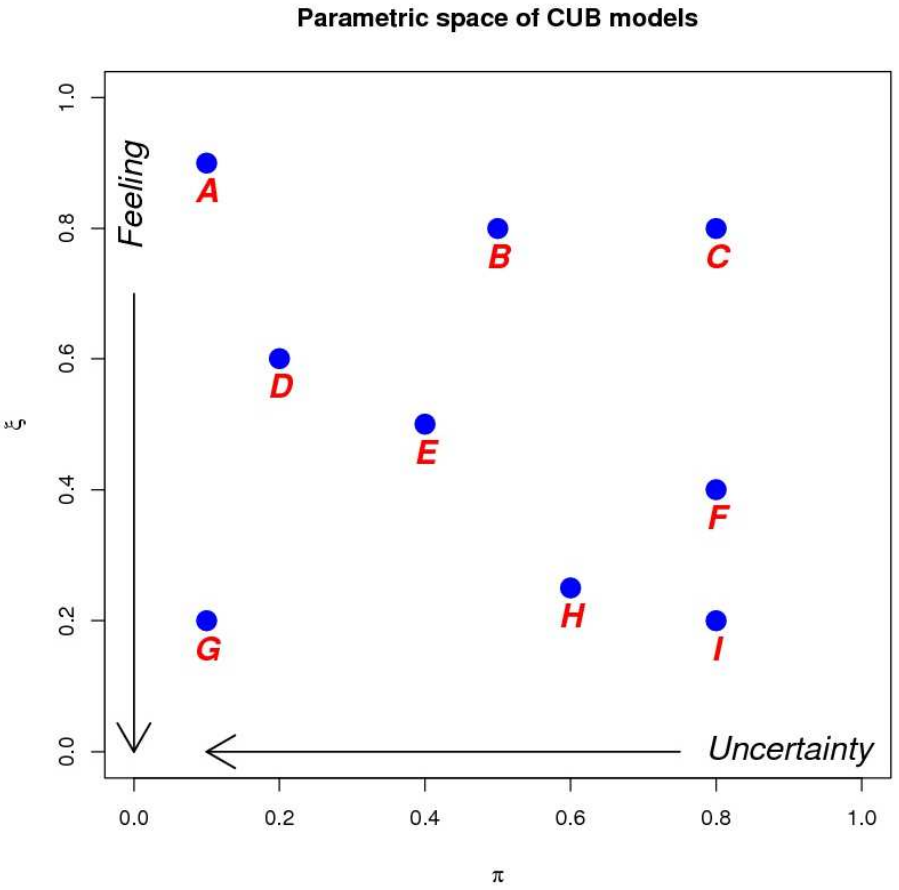
$1 - \pi \rightarrow$ **uncertainty** of the choice

There is a one-to-one correspondence among a CUB random variable and the parameter vector $\theta=(\pi, \xi)'$

$\rightarrow$ each CUB model can be represented as a **point** in the unit square (with coordinates π, ξ)

CUB models


π = 0.1; ξ = 0.9; μ = 4.68
A
π = 0.5; ξ = 0.8; μ = 3.8
B
π = 0.8; ξ = 0.8; μ = 3.08
C
π = 0.2; ξ = 0.6; μ = 4.84
D
π = 0.4; ξ = 0.5; μ = 5
E
π = 0.8; ξ = 0.4; μ = 5.64
F
π = 0.1; ξ = 0.2; μ = 5.24
G
π = 0.6; ξ = 0.25; μ = 6.2
H
π = 0.8; ξ = 0.2; μ = 6.92
I
Prob(R=r)


Parametric space of CUB models
Feeling
Uncertainty
A
B
C
D
E
F
G
H
I
ξ
π

CUB models

CUB models have been extended in several directions (see Iannario & Piccolo, 2011):

- generalizations concerning the **probability distribution** of the components
- inclusion of subjects' and objects' **covariates**
- **multivariate** framework
- **shelter effect** (to model data with an anomalous frequency value on a given response category, Corduas *et al.*, 2009; Iannario, 2012)

CUB models

Inferential issues (specification, estimation, validation) have been obtained also for extended CUB models

A **normalized fitting measure** is

$$\mathcal{F}^2 = 1 - \frac{1}{2}\sum_{r=1}^{m} |f_r - p_r(\hat{\boldsymbol{\theta}})|$$

→ proportion of correct predicted responses

A **program in R** is freely available
(Iannario & Piccolo, 2011)

Case Study

Sensory data coming from a survey on the **Italian espresso**

(Centro Studi Assaggiatori & International Institute of Coffee Tasters)

More than 1,000 Italian and foreign tasters rated (on a 9-point Likert scale) 36 espresso varieties on:

- **Visual** perceptions
  (colour, texture, persistence of the cream)
- **Olfactory** perceptions
  (smell, fragrances or aromas)
- **Gustatory** analysis (flavour, aftertaste)

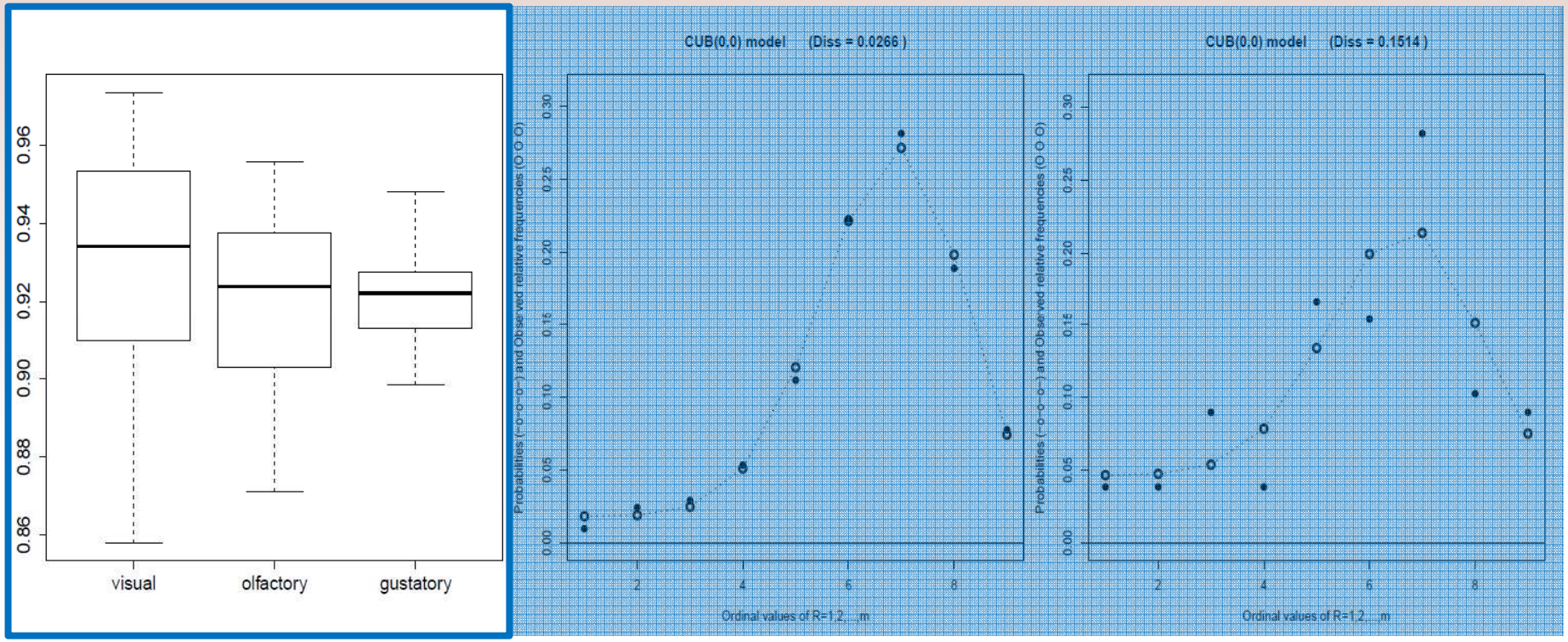
Case Study

We fitted a CUB model for each of the 36 coffees, with respect to the 3 sensory attributes

The estimated models are all **significant** with good fitting measures

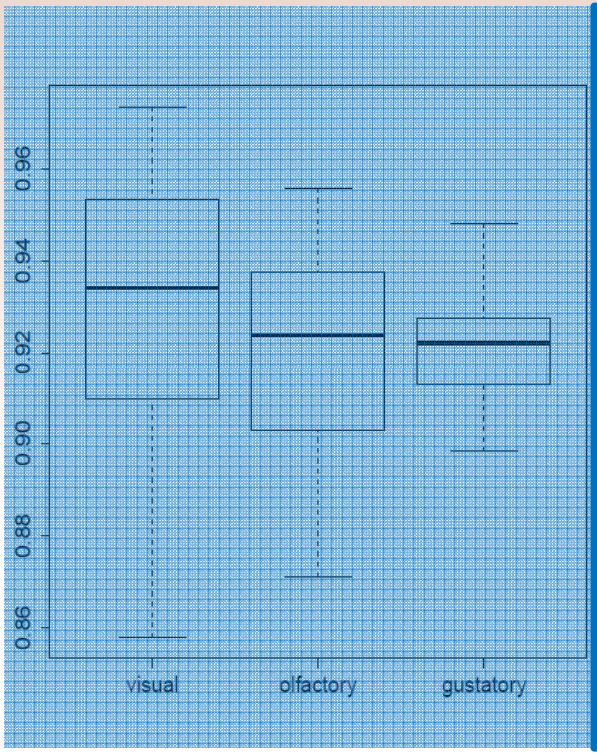
$\mathcal{F}^2$ varies in $(0.849, 0.973)$

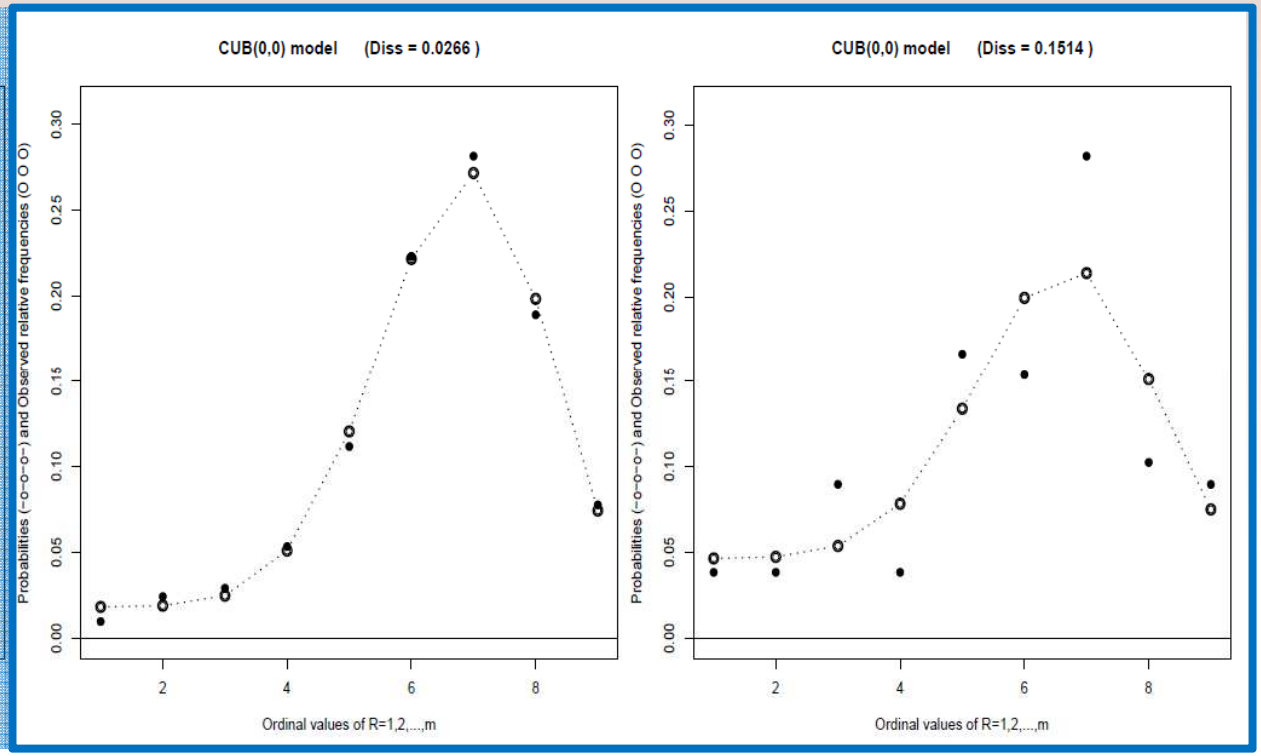
Case Study



Boxplot of $\mathcal{F}^2$ separately for the three sensory attributes

Case Study





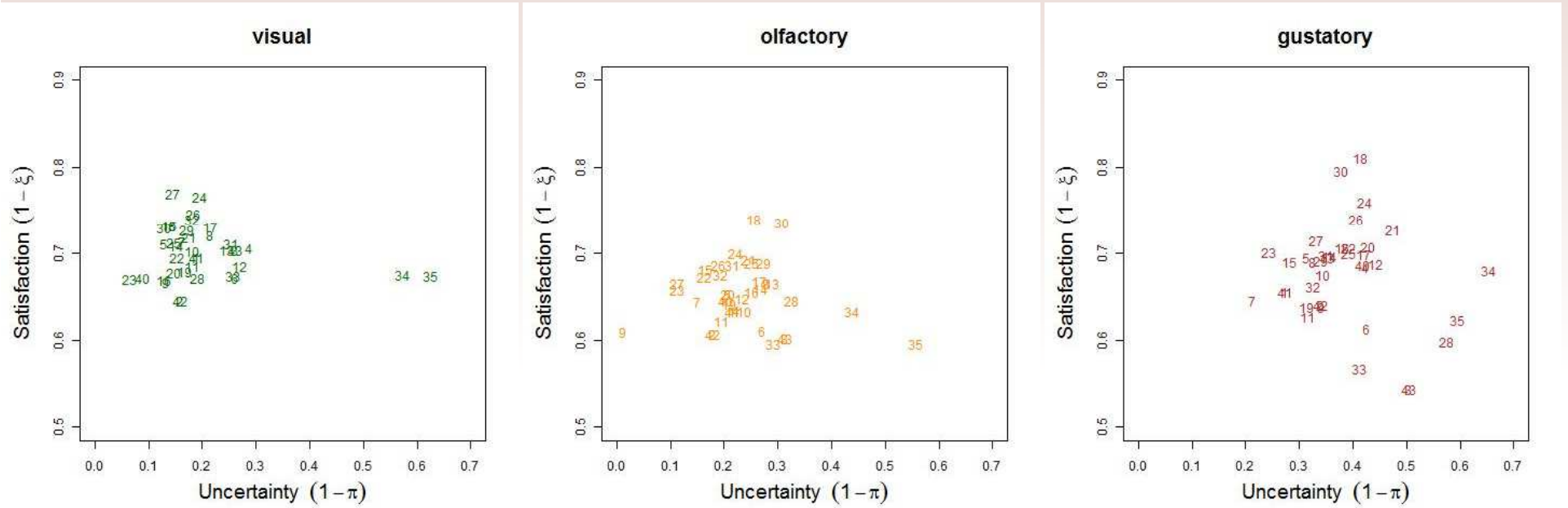
Plot of estimated probabilities *vs* observed relative frequencies

Case Study

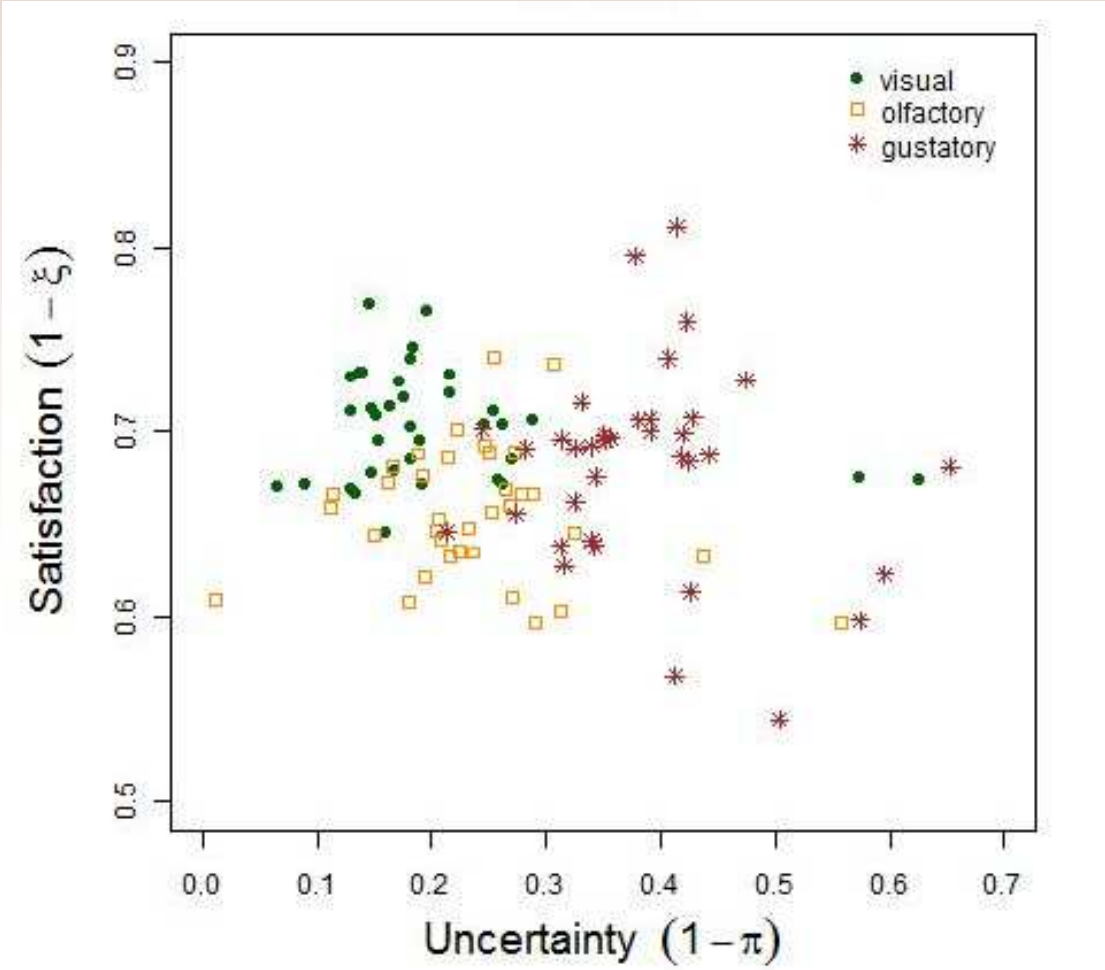
We summarize results by plotting each CUB model in the unit square

This simplifies the complex pattern of the experiment in a **unique representation**

Case Study


visual
olfactory
gustatory
Satisfaction $(1-\xi)$
Uncertainty $(1-\pi)$
0.5
0.6
0.7
0.8
0.9
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Case Study

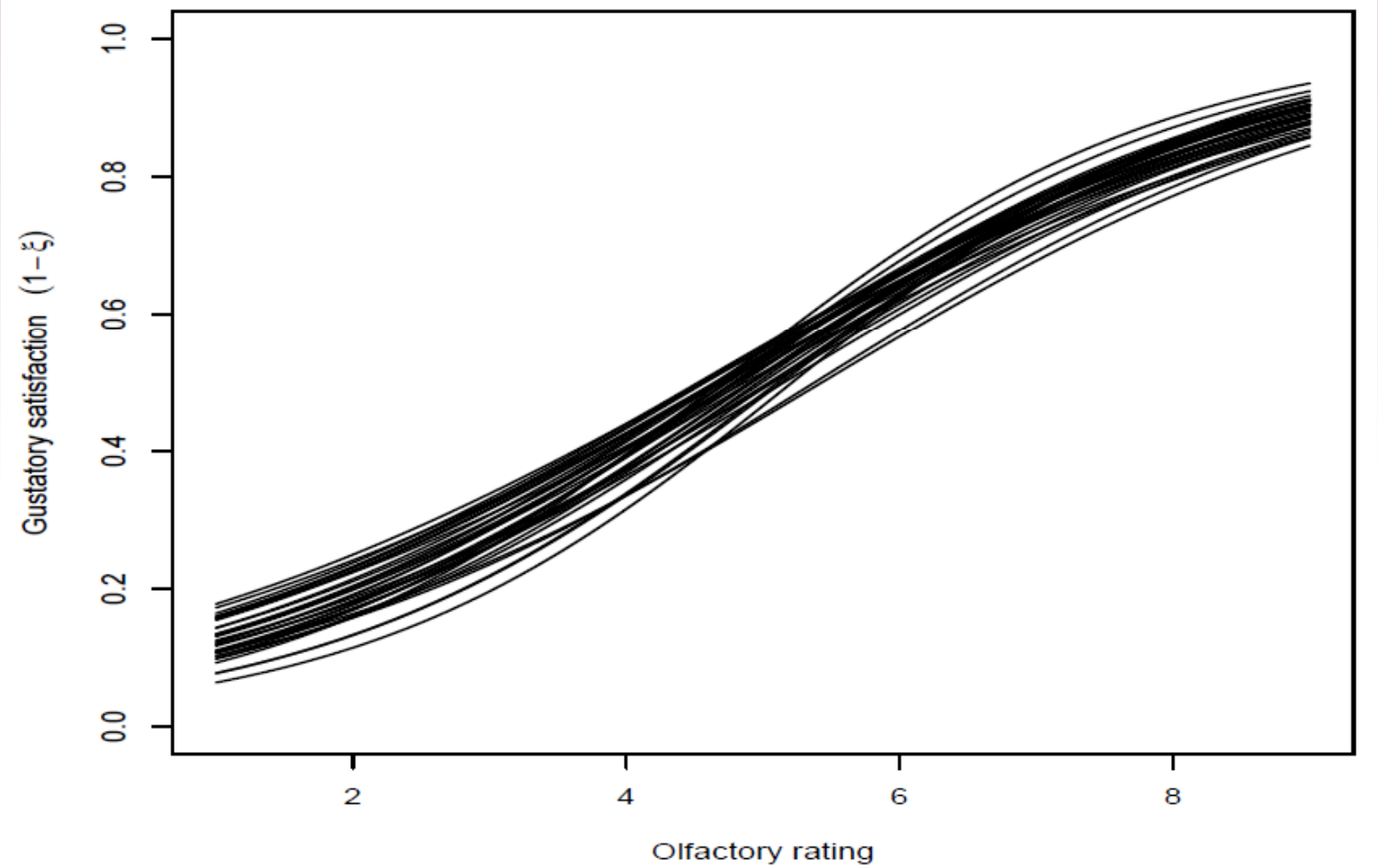

visual
olfactory
gustatory
Satisfaction $(1-\xi)$
Uncertainty $(1-\pi)$
0.5
0.6
0.7
0.8
0.9
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Case Study

An interesting result concerns the relationship between the expressed level of **olfactory rating** and the **gustatory satisfaction** for the 36 coffees, separately

Olfactory scores are used as **covariate** in the CUB models of gustatory satisfaction to see if gustatory satisfaction can be predicted by tasters' olfactory perceptions

Case Study


Gustatory satisfaction (1−ξ)
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
Olfactory rating

Conclusions

CUB models allowed to **interpret** uncertainty and feeling of different coffee varieties and **represent** several results in an effective graphical display

They are useful also for measuring the **predictive ability** of gustatory responses given the olfactory ones

Conclusions

The CUB analysis could be further completed with **product characteristics** in order to better understand relationships among coffees and perceptions and to finally direct manufacturers' efforts to improve their competitiveness

Results from CUB models could be integrated with other **advanced statistical techniques** in order to get a more complete picture of the phenomenon

Main references

Corduas M., Iannario M., Piccolo D. (2009). A class of statistical models for evaluating services and performances, in: M. Bini *et al*. (eds.), *Statistical methods for the evaluation of educational services and quality of products*, Springer, 99-117.

D'Elia A., Piccolo, D. (2005). A mixture model for preference data analysis, Computational Statistics & Data Analysis, **49**, 917-934.

Iannario M. (2007). A statistical approach for modelling urban audit perception surveys, *Quaderni di Statistica*, **9**, 149-172.

Iannario M. (2009). Fitting measures for ordinal data models, *Quaderni di Statistica*, **11**, 39-72.

Iannario, M. (2010). On the identifiability of a mixture model for ordinal data, *Metron*, **LXVIII**, 87-94.

Iannario M. (2012). Modelling *shelter* choices in a class of mixture models for ordinal responses, *Statistical Methods and Applications*, **21**, 1-22.

Iannario M., Manisera M., Piccolo D., Zuccolotto P. (2011). Sensory analysis in the food industry as a tool for marketing decisions, *Submitted for publication*.

Iannario M., Piccolo D. (2009). A program in R for CUB models inference, Version 2.0, available at www.dipstat.unina.it/CUBmodels1/

Main references

Iannario M., Piccolo D. (2011). CUB Models: Statistical Methods and Empirical Evidence, in: Kenett, R.S. and Salini, S. (eds.), Modern Analysis of Customer Surveys, Wiley, NY, 231-254.

Piccolo D. (2003). On the moments of a mixture of uniform and shifted binomial random variables, Quaderni di Statistica, **5**, 85-104.

Piccolo D. (2006). Observed information matrix for MUB models, Quaderni di Statistica, **8**, 33-78.

Piccolo D., D'Elia A. (2008). A new approach for modelling consumers' preferences, Food Quality and Preference, **19**, 247-259.

Piccolo D., Iannario M. (2010). A new approach for modelling consumers' preferences, Proceedings of the 11th European Symposium on Statistical Methods for the Food Industry, University of Sannio, 139-148.

Brentari E., Zuccolotto P. (2011). The impact of chemical and sensory characteristics on the market price of Italian red wines, Electronic Journal of Applied Statistical Analysis, **4**, 265-276.

Brentari E., Levaggi R., Zuccolotto P. (2011). Pricing strategies for Italian red wine, Food Quality and Preference, **22**, 725-732.