e-Agrostat 2021

16th Agro-Industry and Statistical Methods congress

Abstract booklet

Virtual conference, September 14-15 2021

Table of contents

Experimental design

NOVEL STRATEGIES FOR DETERMINING SPARSE COMPONENTS IN PCA AND PLS REGRESSION

El Mostafa Qannari*[1], Essomanda Tchandao Mangamana[2], and Fabien Llobell[3]

[1]Ecole Nationale Vétérinaire, Agroalimentaire et de lálimentation Nantes-Atlantique – Statistiques, Sensométrie, Chimiométrie (StatSC) – Nantes, France
[2]University of Abomey Calavi – University of Abomey-Calavi, 04 BP 1525, Cotonou, Bénin, Benin
[3]Addinsoft – XLSTAT – Paris, France

Abstract

A general strategy for determining sparse components in PCA and PLS regression is outlined. It is simple and intuitively appealing and both algorithms used for PCA and PLS follow exactly the same pattern. It can be easily be extended to other situations and frameworks. Illustrations in sensometrics and chemometrics are discussed and the results are compared to standard methods.

Keywords: Keywords, PCA, PLS regression, sparse methods, sensometrics, chemometrics

*Speaker

Preliminary study of hyperspectral imaging and class-modelling approach for honey authentication

Carolina Silva*[1], Sholeem Griffin[1,2,3,4], Jefferson De Oliveira Mallia[1,5], Juan Attard[1], Owen Falzon[2], and Vasilis Valdramidis†[1,3]

[1]Department of Food Sciences and Nutrition, Centre of Molecular Medicine and Biobanking, University of Malta – MSD 2080, Malta
[2]Centre for Biomedical Cybernetics – University of Malta, Msida, Malta
[3]Centre of Molecular Medicine and Biobanking – University of Malta, Msida, Malta
[4]Metamaterials Unit – University of Malta, Malta
[5]Metamaterials Unit – University of Malta, Msida, Malta

Abstract

Economically motivated adulteration of honey is a common food market issue. Despite the commitment endorsed by the European Parliament of preventing honey counterfeits[1], the Food Fraud summary from the European Commission reported honey products among the most often adulterated food products in the beginning of 2021[2]. Efficient analytical methods and pattern recognition tools can provide an effective way to identify counterfeits and ensure that authentic samples meet EU standards.
Twelve honey samples were produced using different formulations to simulate adulterated honey, with different concentrations of caffeic acid, sucrose and glucose relative to the control honey. Six hyperspectral images of each sample and the control were then acquired at VIS/NIR range (400 to 1000 nm) using the FX10 SisuCHEMA imaging system, with a spectral resolution of 5.5 nm and a lens optics magnification of 0.8. Pre-processing techniques were employed to attenuate unwanted variation in the dataset. Hereafter, an exploratory analysis was carried using principal component analysis (PCA) and, for authentication purposes, two supervised class-modelling techniques were employed and their performance was compared.

After pre-processing, a 2-component PCA model was built, explaining 92% of data variation. The adopted PCA model showed significant differences between the control group and the adulterated samples. The loading plots showed that the differences observed in the score density plot were associated with two spectral regions: the 500-575 nm region, related to riboflavin content, and the 860-1000 nm region, related to protein content. Two class-modelling approaches were then employed for honey authenticity: soft independent modelling by class analogy (SIMCA), and partial least squares - density modelling (PLS-DM). The SIMCA model provided a true positive rate of 92%, but showed poor specificity (78% of false positive for control class). The PLS-DM model using 2 neighbors, smoothing coefficient of 0.4 and autoscaling was built, provided 87% of sensitivity and 95% of specificity for the control class.

*Speaker
†Corresponding author: vasilis.valdramidis@um.edu.mt

The proposed preliminary methodology was found to be promising for screening adulterated honey samples. The results obtained suggest that simpler systems, such as cameras and portable devices can be employed for the purpose of authentication.

https://www.europarl.europa.eu/news/en/press-room/20180226IPR98612/bees-and-beekeepers-meps-set-out-eu-wide-long-term-survival-strategy

https://knowledge4policy.ec.europa.eu/publication/food-fraud-summary-march-2021_en

Keywords: hyperspectral images, honey, chemometrics

Pairing PTR-MS with TDS or TCATA?

Caroline Peltier*[1], Michel Visalli[2], Hélène Labouré[3], Cantin Hélard[1], Isabelle Andriot[2], Sylvie Cordelle[1], Olivier Berdeaux[2], Jean-Luc Le Quéré[4], and Pascal Schlich[2]

[1]Centre des Sciences du Goût et de l'Alimentation [Dijon], Chemosens Platform – Université de Bourgogne, AgroSup Dijon - Institut National Supérieur des Sciences Agronomiques, de lÁlimentation et de lÉnvironnement : UMR6265, Centre National de la Recherche Scientifique : UMR6265, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement : UMR1324, PROBE research infrastructure, $PF_Acronym facility, F-Zip_codeCity, France, AgroSupDijon-InstitutNationalSupérieurdesSciencesAgronomiques, delÁlimentationetdelÉnvironnement : UMR6265--UniversitédeBourgogne-Bât.CSG-9EBdJeannedÁrc-21000Dijon, France$

[2]Centre des Sciences du Goût et de l'Alimentation [Dijon], Chemosens Platform – Université de Bourgogne, AgroSup Dijon - Institut National Supérieur des Sciences Agronomiques, de lÁlimentation et de lÉnvironnement : UMR6265, Centre National de la Recherche Scientifique : UMR6265, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement : UMR1324, PROBE research infrastructure, $PF_Acronym facility, F-Zip_codeCity, France, AgroSupDijon-InstitutNationalSupérieurdesSciencesAgronomiques, delÁlimentationetdelÉnvironnement : UMR6265--UniversitédeBourgogne-Bât.CSG-9EBdJeannedÁrc-21000Dijon, France$

[3]Centre des Sciences du Goût et de l'Alimentation – CNRS : UMR6265, Institut national de la recherche agronomique (INRA) : UMR1324, Université de Bourgone, Agrosup Dijon – Dijon, France

[4]Le Quéré (INRA) – INRA : UMR1324, CNRS : UMR6265, Université de Bourgogne – 17 rue Sully BP86510 21065 Dijon Cedex, France

Abstract

Proton transfer reaction-mass spectrometry (PTR-MS) is an analytical chemistry technique that can be used for measuring over time the concentration of volatile organic compounds directly in the subjects' noses during a tasting. It can be combined with temporal sensory methods such as Temporal Dominance of Sensations (TDS) or Temporal Check All That Apply (TCATA).
This paper aims to compare the conclusions obtained when pairing PTR-MS data with TDS or TCATA. In this purpose, 16 subjects connected to a PTR-MS device evaluated 3 chocolates in TDS and TCATA with 2 replicates for each sensory method.

After relevant pretreatments (breathing correction, removing blank period, normalization...), TDS, TCATA and PTR-MS data were analyzed separately. Then, PTR-MS data was related to sensory data with PLS methods. Thus, the TDS/TCATA durations of sensory attributes were explained by the intensities of relevant ions (independently of products). Bootstrap procedures allowed the stability of the results to be assessed in order to identify potential chemical markers of sensations.
First results showed similar conclusions for TDS and TCATA (same sensory interpretation of the products, same potential biomarkers, ...). PTR-MS data discriminated the products with some potential biomarkers evidenced. The sensory consequence seemed better apprehended by TDS that showed a better product discrimination than TCATA.

*Speaker

Keywords: PTR, MS, TDS, TCATA

TDS or TCATA : product discrimination or panelist repeatability?

Pascal Schlich*†[1], Cindy Frascolla[2], and Michel Visalli[3]

[1]INRAE - Centre des Sciences du Goût et de l'Alimentation – Université de Bourgogne, Institut national de la recherche agronomique (INRA) : UMR1324, CNRS : UMR6265 – Bât. CSG - 9E Bd Jeanne d'Arc - 21000 Dijon, France

[2]Institut de Mathématiques de Bourgogne [Dijon] – Université de Bourgogne, Centre National de la Recherche Scientifique : UMR5584, Université de Franche-Comté, Université Bourgogne Franche-Comté [COMUE] – Université de Bourgogne - 9, avenue Alain Savary - B.P. 47 870 - 21078 Dijon Cedex, France

[3]INRAE - Centre des Sciences du Goût et de l'Alimentation – CSGA Centre des Sciences du Goût et de l'Alimentation, Dijon, France – 9E bd jeanne d'Arc. 21000 Dijon, France

Abstract

Temporal Dominance of Sensations (TDS) and Temporal Check-All-That-Apply (TCATA) are the two more popular methods of temporal sensory analysis. They both aim at characterizing the succession of perceived sensations during the consumption of one intake of a product. In TDS the panelist is asked to select the successive perceived sensations (named as "dominant") in such a way that a single sensation is selected at any time, whereas in TCATA the panelist is asked to select all applicable sensations at any time. In TCATA, the panelist has to deselect a sensation when it disappears, whereas in TDS a new dominant sensation automatically replaces the former one.

TDS has now been used by the sensory community for about 20 years and is supported by a couple of hundreds of publications, whereas TCATA is more recent. However, TCATA is often presented as an improvement of TDS by its advocates. The paper will discuss this topic by using several unpublished data from our lab collected in studies in which both TDS and TCATA were used, sometimes with replicates. Our hypothesis is that generally TDS should offer a better discrimination of the products, whereas TCATA might offer a better repeatability of the panelists. Although, this could be modulated by the type of products under evaluation, the reason for this dichotomy would originated in the difference between selecting one dominant attribute versus all the applicable attributes at a given time.

A new index based on Jaccard dissimilarity and resampling will be used to assess individual repeatability in TDS and TCATA. Behavior of responses in TDS and TCATA will also be compared based on several indicators and finally discrimination will be assessed by multivariate analysis of dominance/applicability durations.

Keywords: TDS, TCATA, Jaccard index, product discrimination, panel performances

*Speaker

†Corresponding author: pascal.schlich@inrae.fr

A modified chi-square framework for the analysis of multiple-response data: application to Free-Comment and Check-All-That-Apply sensory data

Benjamin Mahieu*†[1], Pascal Schlich[1], Michel Visalli[1], and Hervé Cardot[2]

[1]Centre des Sciences du Goût et de l'Alimentation – Université de Bourgogne, Institut national de la recherche agronomique (INRA) : UMR1324, CNRS : UMR6265 – Bât. CSG - 9E Bd Jeanne d'Arc - 21000 Dijon, France

[2]Institut de Mathématiques de Bourgogne [Dijon] – Université de Bourgogne, Centre National de la Recherche Scientifique : UMR5584, Université de Franche-Comté, Université Bourgogne Franche-Comté [COMUE] – Université de Bourgogne - 9, avenue Alain Savary - B.P. 47 870 - 21078 Dijon Cedex, France

Abstract

Introduction

Free-Comment (FC) and Check-All-That-Apply (CATA) are two sensory descriptive methods providing binary data of the same type: presence/absence of several sensory descriptors for each pair of subject and product. At the sensory panel level, these data are summarized into a descriptor by product contingency table. Analyses performed on these data are most often related to the chi-square statistic. However, such practices are not well suited because they consider the citations (one descriptor for one product by one subject) rather than the multiple-responses (vector of citations for one product by one subject) as experimental units. This results in incorrect expected frequencies under the null hypothesis of independence between products and descriptors and thus in an incorrect chi-square statistic. Thus, related analyses, which include Correspondence Analysis, can lead to wrong interpretations.

Methodology

This paper proposes a modified chi-square framework that considers multiple-responses as experimental units. This new framework includes a multiple-response dimensionality test of dependence, a multiple-response Correspondence Analysis, and a multiple-response hypergeometric test to investigate which cells of the contingency table significantly deviate from independence. This multiple-response chi-square framework fits well to FC and CATA data. The outputs from the multiple-response chi-square framework and the usual chi-square framework applied on some CATA data were compared. In these data, an additional artificial product P^* having exactly one-half citation rates for all descriptors of one of the actual products P was added.

Results

Results exhibited that the usual chi-square framework was unable to catch any difference

*Speaker

†Corresponding author: benjamin.mahieu@inrae.fr

between P and P^* while the multiple-response chi-square framework was able to. This demonstrated that when products elicit different citation rates (all descriptors combined), the usual chi-square framework provides counterintuitive and inconsistent outputs while the multiple-response chi-square framework provides logical and consistent ones.

Conclusions
For analyzing multiple-response data, the paper proposes to replace the usual chi-square framework with a new multiple-response chi-square framework taking into account the multiple-response structure of the data. This multiple-response framework is especially relevant to analyze Free-Comment and Check-All-That-Apply sensory data. The new approaches are supported by an R package called "MultiResponseR" that will be demonstrated.

Keywords: Contingency table, Chi square statistic, Correspondence Analysis, Multiple response Correspondence Analysis, Dimensionality test

Title: Optimized time periods segmentation in TDS and TCATA.

Authors: Michel Visalli, Benjamin Mahieu, Pascal Schlich

Temporal Sensation Dominance (TDS) and Temporal Check-All-That-Apply (TCATA) are among the most popular methods of temporal sensory analysis. While TDS asks the panelists to focus on the dominant sensations, TCATA asks them to select all the sensations that apply at all times.

Despite their conceptual differences, standard analyzes of TDS/TCATA data share many commonalities. Curves of attribute proportions of citations is at panel level the closest representation to the raw data, but the interpretation of the continuous time signal can be cumbersome. Alternatives such as ANOVA, PCA or CVA of durations may be easier to interpret, but the sequentiality of perceived sensations is not taken into account. To correct this problem, it has been proposed to divide the continuous time signal into periods, but the choice of the number and the frontiers of periods is most often subjective. Finally, the increase in the number of periods allows maps such as trajectory PCA (TDS) or trajectory CA (TCATA) to be plotted, but the problem of interpretability and thus actionability on the design of the product of such outputs arises again.

The article proposes the “optimized trajectory mrCA” to automatically determine the number and frontiers of periods that best represent the breaks in the temporal perception of products at panel level. The method was on TDS and TCATA data collected from 2 groups of 64 consumers, each consumer participated in 3 separated sessions (1 in lab, two at home) and tasted a total of 20 samples belonging to 4 different product spaces (dark chocolate, guacamole, iced tea and crips). The results show that the “optimized trajectory mrCA” simplifies the interpretation of temporal data and gives insightful information on the key moments of the temporal sensory perception of the products.

Of aircrafts, microbes and statistics. A demonstration of how limitations in data gathering can induce bias in meta-regression models for predictive microbiology

Alberto Garre*[1], Heidy Den Besten[1], and Marcel Zwietering[1]

[1]Wageningen University and Research [Wageningen] – Droevendaalsesteeg 4, 6708 PB Wageningen, Netherlands

Abstract

Introduction: Meta-regression models have gained interest in predictive microbiology during the last years. They are based on the analysis of a large database of microbial attributes (kinetic parameters) gathered from published studies or databases. Most meta-regression models assume an underlying statistical model where the residuals are normally distributed and with constant variance. However, this hypothesis may be violated for microbial inactivation due to experimental limitations. At high temperatures, the inactivation rate may be too high to gather enough data points, biasing the data towards conditions (i.e. strains and media) with relatively high D-values. The opposite happens at low temperatures, since organisms with a high D-value would give no observable inactivation, introducing the opposite bias. We define this bias as "experimental bias".
Methodology: We analysed the datasets used in several meta-analyses already published in the scientific literature for evidence that supports the existence of the experimental bias. Moreover, we used numerical simulations to evaluate its relevance. Finally, we aplied truncated regression (as a generalized non-linear model) to incorporate experimental bias in the meta-regression. It was defined as a Bayesian model to evaluate the evidence supporting the experimental bias of each case studied.

Results: The numerical simulations demonstrate that experimental bias introduces an underestimation of the slope (overestimation of the z-value). This is especially relevant at extreme temperatures (high or low), and it can introduce a bias of up to 1 log-difference in the D-value in the worst case simulated. The Bayesian model supports the hypothesis that there is experimental bias in every dataset studied, although its magnitude varies between them.

Conclusions: There are very good reasons for this experimental bias from the point of view of experimental design. However, it can also introduce a relevant bias in the predictions of meta-regression models. Considering that pasteurization takes place at temperatures higher than those commonly used in experiments, experimental bias can be of high relevance for the design of inactivation treatments. Therefore, it should be accounted for in meta-regression models. In this sense, Bayesian truncated regression can be a tool to correct for this bias without modifying experimental design.

*Speaker

Keywords: robust statistics, experimental bias, Bayesian models, generalized linear models

The Third French Individual and National Food Consumption (INCA3) Survey 2014–2015 and data for microbiological risk assessment: Models for domestic refrigerator temperatures

Juliana De Oliveira Mota*†[1], Hernan Gomez Redondo[2], Hélène Bergis[1], Sonia Poisson[1], Carine Dubuisson[1], Moez Sanaa[1], Virginie Desvignes[1], and Laurent Guillier[1]

[1]Agence nationale de sécurité sanitaire de lálimentation, de lénvironnement et du travail – ANSES - Agence nationale de sécurité sanitaire de lálimentation, de lénvironnement et du travail-0 – 14 rue Pierre et Marie Curie94701 MAISONS-ALFORT Cedex, France

[2]Agence nationale de sécurité sanitaire de lálimentation, de lénvironnement et du travail – ANSES - Agence nationale de sécurité sanitaire de lálimentation, de lénvironnement et du travail-0, Danmarks Tekniske Universitet - DTU (DENMARK) – 14 rue Pierre et Marie Curie94701 MAISONS-ALFORT Cedex, France

Abstract

Introduction: The third individual and national study of food consumption (INCA 3) was carried out to describe consumption habits among the French population between 2014 and 2015. Data were collected about description and quantities of consumed food, beverages and food supplements, but also about the characteristics of households, individuals and their food handlings. The resulting data, such as food storage temperatures at the consumer level, are highly valuable in risk assessment studies.
Methodology: The INCA 3 study was conducted among 5,855 individuals selected according to a three-stage random sample design: geographic units, households, and individuals. The sample included individuals from birth to 79 years of age, living in metropolitan France. Information on food storage temperatures in the refrigerator was collected at the home of the participant, by an interviewer. The survey R package was used to account for the sampling design and weights. The fitdistrplus R package was used to model the variability and uncertainty of temperatures.

Results: A large database was constructed with food storage temperature values in refrigerators, representative of the French population. From 5,855 sampled households, refrigerator temperatures were measured in 5504 households. The observed temperatures were distributed between -3 and 15 Celsius degrees and the median temperature was 6 Celsius degrees. The levels of temperature were described according to the characteristics of the household referent person (e.g. age class, gender), the characteristics of households (e.g. location) as well as the period or season of the collection. The storage temperature distributions were fitted with the most appropriate distribution laws to facilitate Monte-Carlo simulations in the context of microbiological risk assessment models. Finally, to improve

*Speaker
†Corresponding author: juliana.deoliveiramota@anses.fr

the accessibility and reuse of models and data, metadata was described in a harmonized format, following the Minimum Information Required to Annotate food safety Risk Assessment Models principles. The final object was stored on the Risk Assessment modelling and knowledge Integrating Platform to be available for anyone.
Conclusions: A database on food storage temperatures and models in refrigerators from the INCA 3 survey is available in open data and could be useful in the development of future microbiological risk assessment studies.

Keywords: Dietary survey, food storage temperature, fitted models, risk assessment, harmonization

Data-driven modelling for the effect of climate on raw cow milk production traits: the Maltese case study

Lydia Katsini*[1], Carlos André Muñoz López[1], Satyajeet Bhonsale[1], Simen Akkermans[1], Monika Polanska[2], Styliani Roufou[3], Sholeem Griffin[4], Vasilis Valdramidis[4], and Jan Van Impe†[1]

[1]BioTeC+, Chemical Biochemical Process Technology Control, Chemical Engineering Department, KU Leuven – Gebroeders De Smetstraat 1, 9000 Gent, Belgium

[2]BioTeC+, Chemical Biochemical Process Technology Control, Chemical Engineering Department, KU Leuven – Gebroeders De Smetstraat 1, 9000 Gent, Belgium

[3]Department of Food Sciences and Nutrition, University of Malta – MSD 2080, Malta

[4]Department of Food Sciences and Nutrition, Centre of Molecular Medicine and Biobanking, University of Malta – MSD 2080, Malta

Abstract

Climate change is the defining issue of our time and the food systems, as the dairy sector, play a key role because they are both affecting and being affected by climate. In 2018, milk was the second largest agricultural commodity in value terms worldwide, placing significant interest in unraveling the effect of climate on milk production traits. Furthermore, assessing the climatic factors influence on milk production characteristics is essential to evaluate the effects of climate change on the dairy industry. This work aims at determining the effect on both quality and safety milk production traits, i.e., yield, content of protein, fat, and lactose, the Somatic Cell Counts (SCC), and the total microbial load, by considering climatic factors such as temperature, precipitation, wind speed, etc. The case study selected for this work consists of a set of 121 dairy farms located in Malta. Data-driven modelling techniques are exploited, since they are the most suitable tools, given the data available and the complexity of the system. The non-linear dimensionality reduction strategy (t-SNE) was used to facilitate the classification task resulting in the indication that the farms can be allocated to one of the farm-hygienic clusters (i.e., good, medium, and bad hygiene practices), based on the raw milk total microbial plate count levels. PCA results suggest likely relationships between fat content, protein content, and SCC, as well as between lactose content and weekly yield. Finally, PLS models aiming to predict several milk production traits by considering a set of climatic factors were built and validated. In conclusion, despite the high complexity of the underlying phenomena involved, PLS regression models provide reliable predictions in the cases of milk yield and content of fat. By accounting for different likely climate change scenarios, which translates in considering alternate climatic factors future trajectories, exploiting such techniques is valuable to assess the future climate change effects on the dairy industry.

Keywords: Data driven Modelling, Climate Change, Dairy Industry, PCA, PLS, tSNE

*Speaker

†Corresponding author: jan.vanimpe@kuleuven.be

Modelling population growth just became easier with the R package biogrowth and its shiny web application

Alberto Garre*[1], Jeroen Koomen[1], Marcel Zwietering[1], and Heidy Den Besten[1]

[1]Wageningen University and Research [Wageningen] – Droevendaalsesteeg 4, 6708 PB Wageningen, Netherlands

Abstract

Introduction: The analysis of growth is of great interest in many scientific fields. It can indicate the effectiveness of a pesticide, or determine whether a food product is safe to eat. Many mathematical models are available to describe the growth of populations. However, they usually require the application of advanced mathematical techniques, as well as scientific programming. This poses an entry barrier for scientists without a strong background in these specific skills.

Methodology: The functions for modeling population growth have been implemented in the R programming language and compiled as a package openly available from CRAN. It includes functions for fitting models to data gathered under static or dynamic environmental conditions. Moreover, it can be used to predict population growth either as deterministic curves or as prediction intervals, including parameter uncertainty. Calculations are based on several models commonly used in predictive microbiology (Baranyi, modified Gompertz, Logistic, Ratkowsky, Cardinal...). In order to increase the user-friendliness of the software, a web interface to the functions in the software has been implemented using the shiny R package (https://foodmicrowur.shinyapps.io/biogrowth/).

Results: In this presentation will use biogrowth to describe the growth of Listeria monocytogenes under dynamic conditions. We will highlight how fixing some model parameter can improve parameter identifiability, showing also the risks of creating spurious models using this approach. For that, we will use the functions for model fitting and calculation of predictions implemented in the web version of biogrowth, as well as its capabilities to diagnose issues in fitted models.
Conclusions: The functions included in the biogrowth software can make modeling population growth more simple for the scientific community. The fact it is available in two formats increases its user base. Software developers and statisticians can use the functions in the R package to include growth modeling in complex workflows, whereas wet-lab scientists can make use of the advanced modeling tools using the web application without using programing languages. Although the modelling approach is based on predictive microbiology, this modelling approach has also been used to describe other systems. Hence, the software remains interesting for scientists from other fields.

*Speaker

Keywords: kinetic models, R programming, predictive microbiology, user, friendly software

Machine Learning strategies to obtain valency scores from Free JAR data

Alexiane Luc*[1,2,3], Sébastien Lê[3], Mathilde Philippe[2], El Mostafa Qannari[1], and Evelyne Vigneau†[1]

[1]Statistique, Sensométrie et Chimiométrie – Ecole Nationale Vétérinaire, Agroalimentaire et de l'alimentation Nantes-Atlantique, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement : USC1381 – INRAE Site de la Géraudièrerue de la Géraudière BP 7162744316 NANTES CEDEX 3, France

[2]Strategir – Strategir – 5 rue Foy, 33000 BORDEAUX, France

[3]Laboratoire de Mathématiques Appliquées Agrocampus – Agrocampus Ouest, Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement – 2 allée Jacques Frimot - 35000 Rennes, France

Abstract

The Free JAR experiment (Luc et al., 2020) yields two kinds of data: (i) categorization data, which convey hedonic information, and (ii) textual data subjected to have a JAR structure and which highlight the assets and weaknesses of the products. These textual data express a valence towards the products. It seems reasonable to determine valency scores from these opinions, in order to quantify the appreciation implicitly expressed and consider other analyses than those based on textual data. Natural Language Processing methods, and especially sentiment analysis, could help us achieve this goal.
While the lexical approach, based on dictionaries and weights assigned to each word and each grammatical rule, appears rather subjective and time-consuming, we aim to obtain a valency score based on the collected data, with a minimum human intervention. Machine learning algorithms seem well adapted to tackle this issue. In particular, the use of a classifier will allow the machine to understand the link between the hedonic categorization ("I like very much", "I like moderately", "I don't like") and the Free JAR comments. More precisely, the classifier predicts the class to which each comment can be assigned. The probability of a given comment to belong to the class "I like very much" stands as a valency score. Indeed, a high probability expresses that the comment under consideration is unambiguously positive and, contrariwise, a small probability highlights a rather negative assessment.

Different classifiers, including Random Forest (RF) and Support Vector Machine (SVM), are investigated. We will see how the classifier performance can be used to identify a "just about right" level of lexical pre-processing performed on the corpus of comments.
The approaches will be explored using a corpus of free comments elicited from a Free JAR experiment on cheeses. The relevance of the valency scores associated with each comment will be assessed. In particular, they could be compared to the liking scores given by the respondents prior to the Free JAR experiment.

*Speaker

†Corresponding author: evelyne.vigneau@oniris-nantes.fr

Keywords: Free JAR, Machine Learning, valency scores, Natural Language Processing

MULTIVARIATE ANALYSIS OF JUST-ABOUT-RIGHT SCALE DATA WITH OPTIMAL SCALING

Martin Paries*[1], Evelyne Vigneau[1], and Stéphanie Bougeard[2]

[1]Statistique, Sensométrie et Chimiométrie – Ecole Nationale Vétérinaire, Agroalimentaire et de l'alimentation Nantes-Atlantique, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement : USC1381 – INRAE Site de la Géraudièrerue de la Géraudière BP 7162744316 NANTES CEDEX 3, France

[2]ANSES - Laboratoire de Ploufragan-Plouzané-Niort – Agence nationale de sécurité sanitaire de lálimentation, de lénvironnement et du travail – 22440 Ploufragan, France

Abstract

In consumer research, Just-About-Right (JAR) scales are commonly used to identify several key attributes to be modified in a product. Defined as labeled-bipolar scales (e.g. "much too weak", "somewhat too weak", "just about right", "somewhat too strong", "much too strong" for a 5-points scale), JAR scales are intended to combine intensity and acceptability[1]. Beside this inherent complexity, it appears that, (i) the assumption made about the level of scaling of JAR scale (i.e., numerical, nominal, ordinal) is not obvious, (ii) the numerous JAR attributes are rarely analyzed together by a component-based multivariate analysis. Our aim is to propose a descriptive multivariate analysis of sensory attributes measured on JAR scales with respect to their level of scaling. Relationships between JAR attributes and the hedonic assessments are not considered here.
Is JAR scale considered numerical assuming equal space between the scale points? Is JAR scale nominal with, for instance, five categories? In this latter case, categories are considered unordered. Or, is JAR scale of ordinal nature, made of ordered categories? It could be more obvious to consider JAR scale as ordinal but only a limited number of multivariate analysis take into account this level of scaling[1]. Multivariate graphical displays of JAR data are usually obtained from a Principal Component Analysis or a Multiple Correspondence Analysis[1]. In the first case, JAR data are implicitly considered numerical, and in the second nominal.

In this study, we propose to apply Optimal Scaling (OS) as a versatile tool, able to be adapted to the three possible level of scaling. Young[2] introduced a general class of algorithms, denoted Alternating Least Squares algorithm with Optimal Scaling (ALSOS), where Optimal Scaling step alternates with a least squares estimation step of model parameters. Thus, OS makes it possible to consider JAR data in component-based multivariate analysis as numerical, nominal or ordinal. The comparison of the scaling transformation performed and the resulting biplots according to the adopted point of view, will be discussed on a case study on cheese.

*Speaker

REFERENCES

Rothman, L., & Parker, M. (2009). *West Conshohocken, PA: ASTM International.*
Young, F. W. (1981). *Psychometrika.*

Keywords: Just About Right scale, Multivariate Analysis, Optimal Scaling

CATA data: Are there differences in perception?

Fabien Llobell*†[1], Davide Giacalone[2], Sara R. Jaeger[3], and El Mostafa Qannari[4]

[1]Addinsoft – XLSTAT – Paris, France
[2]University of Southern Denmark – Odense, Denmark
[3]The New Zealand Institute for Plant Food Research Limited [Auckland] – Private Bag 92169Auckland Mail CentreAuckland 1142, New Zealand
[4]StatSC, ONIRIS, INRAE – Institut National de la Recherche Agronomique - INRA – Nantes, France

Abstract

Check-All-That-Apply (CATA) experiments are becoming more and more popular for products characterization and development. One of the main objectives of this task is to highlight similarities and differences among the products. CATATIS is a strategy of analysing CATA data that makes it possible to compute an average group configuration of the products taking into account the agreement between the subjects (Llobell *et al.*, 2019). More precisely, the subjects who agree the most with the panel will have a higher weight in the construction of the average group configuration than those who disagree with the panel.
The difference between subjects in product perception is an important topic. While these differences can be highlighted through a clustering approach of the subjects (Llobell *et al.*, 2019), the problem of assessing differences in perception between predefined groups of subjects has not been sufficiently investigated. For example, 'are the products differently perceived by men and women?'

In order to address this issue, we propose a permutation test that assesses whether the difference between the two groups is significant. The test draws from a hypothesis testing strategy proposed by Meyners (2001) to assess differences in liking between two groups.

The test statistic is based on the Ochiai coefficient between the two CATATIS group average configurations. We randomly permute the subjects in the two groups by keeping their sizes unchanged. For each permutation, we compute the Ochiai coefficient between the two CATATIS group average configurations of the permuted groups. The proportion of values of the test statistic that are lower than the observed value yields a pseudo p-value of this test (Jakobowicz, 2007).

The general strategy of analysis is illustrated based on a case study focusing on product perception of beverages.

References

Jakobowicz, E. (2007). *Contributions aux modèles d'équations structurelles à variables latentes* (Doctoral dissertation).

*Speaker
†Corresponding author: fllobell@xlstat.com

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *FQ&P*, *72*, 31-39.
Meyners, M. (2001). Permutation tests: Are there differences in product liking?. *FQ&P*, *12*(5-7), 345-351.

Keywords: CATA, Test, Differences

Less data, same relevance: optimizing paired comparisons sensory evaluation

Julien Rogues*[1] and Emira Mehinagic*[1]

[1]SPF diana – Diana Pet Food Business – zone du Gohelis, France

Abstract

Paired comparisons are increasingly used in sensory evaluation. The Bradley-Terry-Luce model is the most common to analyze paired comparison data [1]. However, the simplicity of this method is counterbalanced by the high number of pairs that each subject has to evaluate to obtain robust results.
Several approaches may be used to optimize paired comparisons sensory evaluation without decreasing the statistical relevance of the analysis. This study presents the impact of different optimization strategies on product ranking when applied on paired comparison data from pet food sensory evaluation made by pets.

Following the approach of Westland and Burton [2;3], several simulation studies were created from more than 80 complete paired comparison design (from 4 to 17 products) conducted on dogs and cats at Panelis, expert center in palatability measurement. Two strategies were chosen to reduce the number of data involved in the Bradley Terry model:

1 - Reduction of the number of animals involved in each paired comparison test

2 - Reduction of the number of pairs tested, with 2 options:

2a - removing randomly a certain number of pairs of products

2b - removing specific pairs: the pair with the highest and the lowest palatable products, some pairs of products belonging to the same group of preference...

This last option was done using preliminary information about the expected ranking of the studied products.

These different options were compared based on the correlation coefficient and the Euclidean distance between the coefficients calculated from the Bradley-Terry model on complete data and subset data.
Results showed various cases in which the reduction of data, either by decreasing the number of animals or the number of pairs, had a small effect on the final ranking of the products. The new method could allow sensory scientists increasing the number of products compared while keeping a relatively low number of tests. This will enable both cost and time savings.

Keywords: data reduction, sustanaible statistics, paired comparisons, bradley Terry, petfood

*Speaker

How statisticians can help dogs shape their summer body: validation of a new method to assess dog food satiating performance

Céline Baron*[1], Anne Lepoudere[1], and Emira Mehinagic[1]

[1]Diana Pet Food – Diana Pet Food – Elven, France

Abstract

Dogs' overweight and obesity issues have noticeably increased during the last decades. Excessive body fat accumulation is a serious health concern that affects dog's quality of life. Different nutritional strategies exist to avoid weight excess. One of them consists in limiting food intake by using satiating ingredients in the diet formula.

Satiety gathers all signals elicited by food ingestion that contribute to develop fullness sensation and to inhibit hunger. Since dogs can't talk, evaluating their level of satiety is a real challenge.

The aim of this study was to develop a reliable, non invasive method to evaluate the level of satiety induced by different dog foods.

3 diets were formulated: a control diet and the same diet enriched with 2 doses of protein hydrolysates used as satiating ingredients.

In a preliminary trial, the palatability of the 3 diets was evaluated with expert panel of 40 dogs using 3 versus tests.

A crossover study was then built: 32 expert dogs were fed with the 3 diets in a randomized order. Each diet was presented to dogs during 2 days with 2 meals per day. A commercial dog diet - challenge test diet - was introduced 40 min after the last meal of the 2nd day. The hypothesis was that hormonal signals related to satiety act within 40 minutes after ingestion.

Results of the preliminary trial showed that the enriched diets were more palatable than the control. However, consumption decreased significantly the second day for these diets contrary to the control. Moreover there was a significantly lower consumption of the challenge test diet when dogs were fed with diets containing hydrolysates. The method allowed highlighting the satiety effect of the hydrolysates.
This newly developed "challenge test" appeared to be an accurate method to evaluate dog food satiety. Its efficacy needs to be confirmed with other types of satiating ingredient and It could be upgraded to get kinetics of satiety by testing different periods of time for the challenge test diet presentation.

Keywords: Challenge test, crossover, satiety, dogs, method developpement

*Speaker

An original application of WSP algorithm to select the best extra virgin olive oils awarded in an international contest

Fabien Girard[1], Jacques Artaud[1], Christian Pinatel[2], Catherine Rebufa[1], and Magalie Claeys-Bruno*[1]

[1]Aix-Marseille Université – Aix Marseille Univ, Univ Avignon, CNRS, IRD, IMBE, Marseille, France – Site de l'Etoile Avenue Escadrille Normandie Niémen 13013 Marseille, France

[2]Centre Technique de l'Olivier – Centre Technique de l'Olivier – Maison des Agriculteur 13100 Aix-en-Provence, France

Abstract

Extra virgin olive oil is the one of the only agri-food products for which the physical, chemical and organoleptic characteristics are regulated by the European Union, the International Olive Council and the Codex Alimentarius depending on where they are trade. Many international olive oil contests are organized in different countries to promote olive oil by being the meeting places between jurors, producers and consumers for exchanges on social, technical and commercial aspects. Their regulation is generally based on the rules where olive oil samples are often selected according to the nature of their fruitiness (green or ripe) and the intensity of their fruitiness (light, medium or robust). But this categorization penalizes the producers when they must allocate *a priori* their samples in one category without reference. The Agency for the Valorization of Agricultural Products (AVPA, France) classes the sample in six clusters at the final step of selection. Also, the organoleptic properties of olive oil samples are evaluated from tree specific descriptors: (i) the fruitiness obtained by the combination of the intensity, the harmony of the fruitiness and the intensity of the maximum defect, (ii) the structure parameter obtained by weighting the importance of the intensity of bitterness and that of the intensity of pungency to take into account the consumer preferences, (iii) the aromatic maturity, a specific descriptor representing the fruitiness maturity (green or ripe). In order to select 30% of olive oil samples at the final stage, AVPA methodology is to place olive oil samples in 2D space (structure versus aromaticity) to define the fruitiness categorization: green (maturity score $<$ 0.5), ripe (maturity score $>$ 0.5) and the fruitiness intensity (light, medium and robust) in according with a range of structure descriptors, different from year to year to account for variable agroclimatic conditions. Here, we proposed the use of WSP algorithm to select olive oil samples in this 2D descriptors space in order to have a representative selection. The choice of an appropriate minimal distance between two points allowed to select the best olive oils locally even if they were placed on the border of the categories.

Keywords: WSP algorithm, olive oil, selection of point, descriptors

*Speaker

Effect of Four Hydrocolloids on Textural Quality Attributes of Gluten-Free Batter and Bread

Christian R. Encina-Zelada*†[1,2,3], Vasco Cadavez‡[1], José A. Teixeira[3], and Ursula Gonzales-Barron§[1]

[1]CIMO Mountain Research Centre, School of Agriculture, Polytechnic Institute of Braganza (IPB), Portugal – Campus de Santa Apolónia, 5300-253 Bragança - Portugal, Portugal
[2]Departamento de Tecnología de Alimentos, Facultad de Industrias Alimentarias, Universidad Nacional Agraria La Molina, Peru – Av. La Molina s/n, Lima12, Lima, Perú, Peru
[3]Centre of Biological Engineering, School of Engineering, University of Minho, Portugal – Campus de Gualtar 4710-057 Braga Portugal, Portugal

Abstract

This study aimed to compare the separate effects of xanthan gum (XG), guar gum (GG), hydroxypropyl methyl cellulose (HPMC) and carboxymethyl cellulose (CMC) on gluten-free batter and bread quality; and to explore the interrelationships among 20 measurements of physicochemical and rheological attributes of batter and bread. Thirty-six formulations were produced with 1.5, 2.5, 3.5% XG and 90, 100, 110% water (9 combinations); 2.5, 3.0, 3.5% GG and 90, 100, 110% water (9 combinations); 3.0, 4.0, 5.0% of HPMC and 80, 90, 100% water (9 combinations); and 3.0, 4.0, 5.0% of CMC and 80, 90, 100% water (9 combinations). A principal component (PC) analysis revealed that the information contained in the 20 variables could be effectively decomposed into three major components, firstly related to bread crumb porosity, baking loss and hardness and batter stickiness (PC1=44.1%), secondly linked to batter firmness and strength-cohesiveness (PC2=16.4%), and lastly bread crumb springiness and cohesiveness (PC3=11.2%). Higher doses of XG, GG, HPMC or CMC hydrocolloids produce less sticky batters, but of increased firmness. Higher doses of XG, GG or CMC gums produce loaves of reduced specific volume with a more cohesive texture and sliced bread of smaller pores; while higher doses of HPMC gum produce baked loaves of greater specific volume with a more cohesive, less springy and a denser crumb texture.

Keywords: Principal component analysis, gums, xanthan, guar, HPMC, CMC, batter rheology, crumb texture.

*Speaker
†Corresponding author: cencina@lamolina.edu.pe
‡Corresponding author: vcadavez@ipb.pt
§Corresponding author: ubarron@ipb.pt

Effect of partial substitution of wheat flour with lupine (Lupinus mutabillis) and kañiwa (Chenopodium pallidicaule) flours on cookie quality

Álvaro J. Bracamonte-Herrera*[1], Dario M. Cabezas*†[2], María Jimena Correa*‡[3], Sylvia Alcázar-Alay[4], Julio Vidaurre-Ruiz[4], Ritva Repo-Carrasco-Valencia*§[4], and Christian R. Encina-Zelada*¶[4]

[1]CIINCA (Centro de Investigación e Innovación en Productos Derivados de Cultivos Andinos), Facultad de Industrias Alimentarias, Universidad Nacional Agraria La Molina – Av. La Molina s/n, La Molina 15024, Lima, Perú., Peru

[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) – Godoy Cruz 2290, Ciudad Autónoma de Buenos Aires, Argentina

[3]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) – Godoy Cruz 2290, Ciudad Autónoma de Buenos Aires, Argentina., Argentina

[4]CIINCA (Centro de Investigación e Innovación en Productos Derivados de Cultivos Andinos), Facultad de Industrias Alimentarias, Universidad Nacional Agraria La Molina – Av. La Molina s/n, La Molina 15024, Lima, Perú., Peru

Abstract

The global cookie market is in constant reformulation and reassessment. This behavior has primarily been driven by changing consumer expectations. Lupine (*Lupinus mutabillis*), and kañiwa (*Chenopodium pallidicaule*) are native food crops grown in the Andean region and used as food by the Incas and previous cultures. The nutritional value of these plants is related to the high protein, dietary fiber, mineral, and vitamin content of the seeds. The main objective of this work was to evaluate the effect of the partial substitution of wheat flour with lupine (*Lupinus mutabillis*) and kañiwa (*Chenopodium pallidicaule*) flours on cookie quality. A mixture design approach was used to determine the interaction effects of the three flours with the following content restrictions: wheat 40-70% (WF), lupine 10-20% (LF), and kañiwa 20-40% (KF). The response variables were the spread factor (SF, shape characteristic), the fracture point (FP, texture characteristic), and the protein content (PC, nutritional characteristic) of the cookies. Formulations with higher wheat flour content showed greater SF values and resistance to fracture (high FP values). These results are related to gluten proteins and microstructure characteristics that gluten confers to the dough matrix. On the other hand, the formulations with higher content of Andean crops flours generated higher SF values, followed by an increased content of great nutritional value proteins. Cookies

*Speaker

†Corresponding author: 77.dario@gmail.com

‡Corresponding author: jimecorrea@hotmail.com

§Corresponding author: ritva@lamolina.edu.pe

¶Corresponding author: cencina@lamolina.edu.pe

showed an adequate shape and nutritional benefits of the Andean flours instead of wheat. The optimum cookie formulation was 40% WF, 20% LF and 40% KF, getting the highest desirability value (0.778). These results show the aptitude of kañiwa and lupine flours to be used to develop cookies with acceptable technological features, high replacement values, and improved nutritional qualities.

Keywords: Mixture design, Andean crops, Desirability, Optimization

Optimization of a photocatalytic water treatment using response surface methodology and desirability function

El Mountassir El Mouchtari[1,2], Magalie Claeys-Bruno*[3], Salah Rafqah[1], Hafid Anane[1], Anne Piram[2], Samir Briche[4], and Pascal Wong-Wah-Chung[2]

[1]Université Cadi Ayyad – LCAM, Faculté Polydisciplinaire de Safi, Morocco
[2]Aix-Marseille Université – Laboratoire Chimie Environnement (LCE) UMR 7376 – Europôle de l'Arbois- Bât. Villemin BP 80 13545 Aix-en-Provence Cedex 4, France
[3]Aix-Marseille Université – Aix Marseille Univ, Univ Avignon, CNRS, IRD, IMBE, Marseille, France – Site de l'Etoile Avenue Escadrille Normandie Niémen 13013 Marseille, France
[4]Département stockage de l'énergie et revêtements multifonctionnels – MAScIR Rabat, Morocco

Abstract

Even if many previous works in the literature focused on the elimination of pharmaceutical pollutants in water by heterogeneous photocatalysis with good efficiency, too few of them were interested in the simultaneous determination of the major parameters controlling the adsorption and induced degradation process. Mostly, the effect of one or two parameters was measured on the removal from water, making it impossible to say whether the optimal experimental conditions for adsorption and induced degradation have been reached. However, some studies pointed out the effect of some parameters. Design of experiments (DOE) methodology has been applied to optimize two criteria: adsorption and photodegradation capacities to eliminate carbamazepine (CBZ) in aqueous solution in the presence of a different composite material under light exposure. A non-classical experimental design was performed considering two quantitative factors (pH, and temperature) and two qualitative factors (oxygen concentration with 3 levels and nature of the material with 7 levels). A second order multiplicative model with interaction effects between qualitative and quantitative variables has been considered and 66 experiments have been performed to calculate the coefficients. From the response surface methodology (RSM), the optimized operational factors for CBZ adsorption and degradation were independently determined. Using desirability function approach, common optimized conditions for both responses were assessed.

Keywords: Optimization, Desirability, Water treatmet

*Speaker