Content is avaliable at SCOPUS

Meat Technology — Special Issue 64/2

www.meatcon.rs ▪ www.journalmeattechnology.com

UDK: 637.5.053

ID: 126133513

https://doi.org/10.18485/meattech.2023.64.2.7

Review paper

(Mis)understandings in research methodology and chemometrics in meat science

Predrag Putnik[a], Daniela Šojić Merkulov[b], Branimir Pavlić[c], Branko Velebit[d] and Danijela Bursać Kovačević[e]*

[a] *Department of Food Technology, University North, Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia*
[b] *University of Novi Sad Faculty of Sciences, Department of Chemistry, Biochemistry and Environmental Protection, Trg Dositeja Obradovića 3, 21000 Novi Sad, Serbia*
[c] *Faculty of Technology, University of Novi Sad, Blvd. Cara Lazara 1, 21000 Novi Sad, Serbia*
[d] *Institute of Meat Hygiene and Technology, Kaćanskog 13, 11040 Belgrade, Serbia*
[e] *Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, 10000 Zagreb, Croatia*

ARTICLE INFO

Keywords:
Research methodology
Data analysis
Mathematical modeling
Principal component analysis
Hypothesis testing
Chemometrics
Meat production

ABSTRACT

The main purpose of this manuscript is to discuss the most common errors in reporting food science data, with special attention to meat science, while offering suggestions that are common and long known in the regular research methodology of any field. Because the quality of data determines the quality of conclusions that are decisive for subsequent actions and the allocation of (often scarce) resources, low quality data can be a barrier to progress in the field rather than paving the way to a better understanding of the important aspects of food (meat) production. For valid conclusions, it is important to define hypotheses for a particular data collection, to collect data correctly, and to choose the right test for analysis. If a professional in meat production needs to optimize or predict a particular production outcome, mathematical modeling is the right choice. On the other hand, if one is looking for structure within the data, principal component analysis (PCA) is one of the valid options. Both approaches have unlimited applications in meat and food science in general, which can also provide various benefits for industrial purposes, such as getting ahead of competitors in the market (by identifying optimal customers, predicting customer acceptance of a meat product, various aspects of business intelligence such as improving effectiveness and efficiency etc.).

1. Introduction

All branches of science have the same common denominator, namely the generation of data and the need to analyze these data in order to draw valid conclusions (*Green et al*., 2007). This is accentuated by the accelerated development of instrumental analysis in modern wet chemistry, which is capable of generating large datasets from routine analyzes (*Varmuza & Filzmoser*, 2009). For instance, it is now quite common for analytical methods, such as gas chromatography, to yield datasets containing information on hundreds of compounds in meat products (*Sohail et al*., 2022), which can be difficult to understand unless the research team has experienced data analysts.

However, due to necessity and need, the majority of time, statistical methods for data analysis are superficially learned as some side skills by researchers who are originally experts in different fields of

***Corresponding author:** Predrag Putnik, *pputnik@alumni.uconn.edu*

Paper received August 9th 2023. Paper accepted August 30th 2023.
Published by Institute of Meat Hygiene and Technology — Belgrade, Serbia

research. Unfortunately, that outcomes with number of shortcomings in research and questionable usefulness of the data obtained (*Ali & Bhaskar*, 2016). Egregiously, the conclusions drawn then do not best serve their intended purposes. We have reached the point in food science that quite often not even basic notations are done correctly.

Here, the first error that comes to mind is the way of denoting the probability in majority of statistical tests, the well-know and used p-value (i.e., the probability calculated by a statistical test) that is commonly and wrongly reported in food science papers as 'p<0.05' rather than 'p≤0.05', which is the correct way to refer to this statistical metric. The former notation (use of the character 'less-than') symbolizes that 0.05 is not the limit of significance (anything below this value is significant, but not the sole 0.05 value), whereas the later notation ('less than or equal to') includes 0.05 as a value that properly outlines the significance of the statistical test.

Numerous studies do not even contain a correctly written hypothesis, and seriously lack experimental design or examination of the suitability of different statistical tests for different datasets, which will be discussed later in the text. The purpose of this paper is to draw attention to the most common errors in the reporting of food science data, with particular emphasis on meat science, while providing the correct approaches that are a common part of research methodology and have been established for decades, well back into the last century.

2. Materials and methods

The data in this manuscript were obtained from common sources such as Web of Science, PubMed, Scopus, and other platforms that provide scientific references. In addition, various search engines such as Google.com, Ask.com, Bing.com, and other, were used to search for relevant terms. The concept of the paper was designed to include selected examples of research methodology, from common basic (mis)understandings to complex ones with a focus on meat science. It is important to note that this short manuscript is far from sufficient to list all methodological issues encountered in meat science. Rather, it should be seen as the 'tip of the iceberg,' encouraging readers to investigate further and fill in the gaps in their own knowledge of research methodology, so the entire food science field provides information of the highest possible quality.

3. Hypothesis testing and why it is used

Most researchers in food science are familiar with the concepts of stating hypothesis and that all research should include some form of hypotheses. What is less known is that the purpose of hypotheses is to test them with some statistical test with purpose of rejecting or accepting either the null hypothesis or hypothesized alternative. A null hypothesis generally conveys absence of difference, while an alternative hypothesis states the presence of difference in the research (usually among studied groups).

However, it is frequently observed that in doctoral dissertations, research papers and various research proposals, the research hypotheses are listed in an erroneous way. Usually, this is done in a descriptive manner in which the authors explain what they hope to achieve from their research. For example, a commonly stated hypothesis may sound like this:

'*Selected meat products with a high protein content that are commonly present in them will increase the levels of amino acids in the diet*'

In the above example, there are the following issues: first it is not known what kind of hypothesis is being talked about. Second, the hypotheses should be stated at least in pairs (with at least the null hypothesis H_0 paired with an alternative hypothesis or H_1), and third, they should be stated for a particular data analytic test. Another problem with the above hypothesis is the specification of direction. Saying '*…content will increase…*' assuming that the data are normally distributed, means that only half of the normally distributed data will be analyzed. The correct formulation of the hypothesis in meat science for the above example and for normally distributed data to be analyzed with the independent *t*-test should be something like this:

'*H_0: $\mu_1 = \mu_2$ or in plain English: meat products with a high protein content from group 1 will yield* ***equal*** *average levels (means) of amino acids (g/100 g product) in the diet as meat products with a high protein content from group 2'*

and

'H_1: $\mu_1 \neq \mu_2$ or in plain English: meat products with a high protein content from group 1 will yield ***different*** *average levels (means) of amino acids (g/100 g product) in the diet from meat products with a high protein content from group 2'*

In stating hypotheses as the above example, all methodological requirements were met and readers

knows what the authors hypothetically compared (i.e. quantities of amino acids among two different groups of meat products expressed as grams of amino acids per 100 grams of meat product).

4. Use of experimental design and why it is needed

Experimental design is the most important step to obtain valid data and conclusions (*Croarkin*, 2013). It consists of defining important experimental factors that are under the control of the experimenter (also referred to as independent variables) and linking them to response variables, correspondingly referred to as dependent variables, so that they can be modeled for the purpose of extrapolation (prediction). The experimental design is usually given in a table (e.g., Table 1) and is used by analytical chemists as a navigational map when conducting series of experimentations (e.g., runs). In other words, it specifies the order of experimentations and the combination of independent variables expected to alter the dependent variables (Figure 1). Experiments should be in randomized order to prevent any potential biases, i.e., to provide homogeneity of variance in the examined groups.

Table 1. Example of experimental design table

Experimental run	Replications	Pressure	Temperature	Time	Microbial load	Amino acid content	Vit. B_{12} content
		(MPa)	(°C)	(min)	CFU	g/100 g meat	µg/100 g
1	1	10	30.0	15.0			
2	1	100	15.0	5.0			
3	2	10	15.0	15.0			
4	2	100	30.0	5.0			
5	3	100	15.0	15.0			
6	3	10	30.0	5.0			
7	4	10	15.0	5.0			
8	4	100	30.0	15.0			
9	5	10	30.0	15.0			
10	5	100	15.0	5.0			
11	6	10	15.0	15.0			
12	6	100	30.0	5.0			
13	7	100	15.0	15.0			
14	7	10	30.0	5.0			
15	8	10	15.0	5.0			
16	8	100	30.0	15.0			

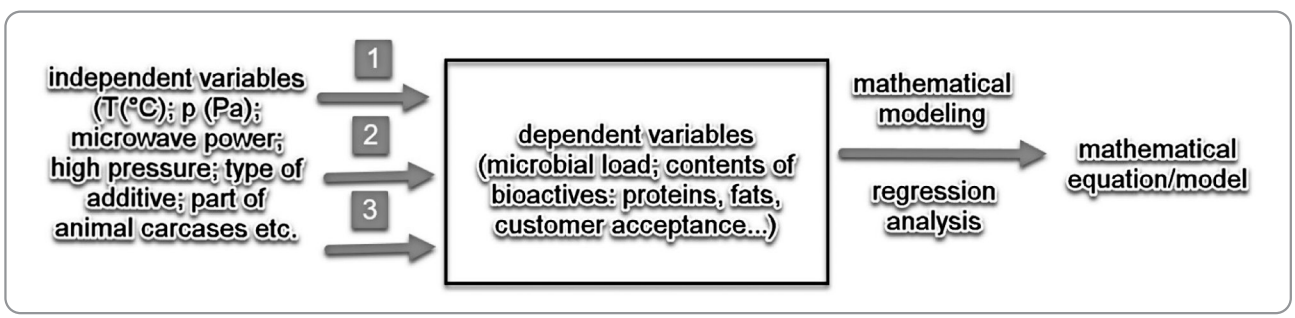

Figure 1. Summary of mathematical modeling.

Dependent and independent variables can be grouped as qualitative or quantitative (*Szymańska et al*., 2015), with the former including nominal variables (e.g., meat types, meat seasonings), dichotomous variables (authentic/adulterated meat products, male/female animals etc.), and ordinal variables (the data have some rankings, as levels of the hedonic scale for sensory evaluation of cured sausage). In contrast, quantitative variables include scales, intervals, and ratios (*Larson-Hall*, 2010). In meat science, examples of scales include the duration of processing in minutes, the temperature of bratwurst, the amount of salt added to the brine, etc. Intervals and ratios are usually similar to terms used in everyday life.

5. Use of mathematical modelling and principal component analysis

In the field of food science, engineers often try to predict or optimize certain aspects of food production (*Croarkin*, 2013). Usually, this means that for a particular process, a particular dependent variable is as high/low as possible. In the meat industry, for example, market pressures dictate that products must have the highest possible microbial safety, while minimizing the use of (unpopular) additives such as nitrates and nitrites (*Šojić et al*., 2022). Therefore, in order to optimize (in this case, minimize) the use of additives for a preservation process (e.g., high-pressure processing), we construct mathematical equations that are capable of describing, to a meaningful (significant) extent, the changes in microbial load upon addition of that particular additive while simultaneously accounting for the changes in high-pressure parameters. If constructed according to valid methodological principles, created mathematical equation should be able to follow (predict) changes across the entire spectrum of variations in high-pressure parameters while simultaneously allowing for altering amounts of the unpopular additive, and perhaps finding the point at which high-pressure processing is sufficient to provide full microbial safety without the use of nitrates/nitrites, even if this is in settings that were not originally included by experimental design. This process is called extrapolation, and the predictions can later be tested in the laboratory to observe the full accuracy of a model.

Here it is important to note that mathematical models are usually built by some type of regression analysis, using potentially relevant factors for testing that were initially chosen by experts in the field of meat science (production technologists, engineers etc.). Clearly, the relevance of each factor should be tested for significance, and all insignificant components of a model must be omitted from the equation as they serve no purpose. This is typically achieved by some multivariate analysis of variance (MANOVA). In practice, optimization/prediction for industrial purposes can provide various benefits, such as an edge over competitors (e.g., efficient production of food, optimal exploitation of the market, identification of customers from the existing customer pool, prediction of customer acceptance, various aspects of business intelligence). Another possibility is to embed models in computer software for commercialization or public access (*Putnik et al*., 2017).

Another data analysis tool commonly used in meat science comes from the group of factor analyses known as principal component analysis (PCA) (*Granato et al*., 2018). This type of analytical approach is also known as ‘dimension reduction’ because the main idea is to reduce a large number of correlated variables from the dataset by using as few uncorrelated factors as possible. The main purposes of PCA are to find underlying correlation patterns within a dataset or to identify structural patterns, i.e., to create indices, otherwise known as principal components or factors (hence the name, PCA).

Often, PCA is erroneously used to group data according to certain criteria, which is not the purpose of this tool; this data grouping should be avoided due to its questionable usefulness. For the purposes of grouping (clustering), there is an entire range of cluster analyses that differ from PCA (*Aldenderfer and Blashfield,* 1984). Moreover, a large number of PCA studies in scientific papers do not even check whether the given data are suitable for this analysis. For example, it is very rare to find statistical metrics for the Kaiser-Meyer-Olkin (KMO) test and the Bartlett’s test for sphericity in meat science reports; these two tests are needed to check that the basic requirements are fulfilled for performing PCA. The KMO test measures the proportion of variance in a dataset that might be caused by the underlying factors, while Bartlett’s test verifies whether the initial variables in a dataset are correlated (*Tabachnick and Fidell*, 2007). Both of these tests should be reported together with complete factor loadings for all factor analyses, including for PCA.

6. Chemometrics and meat production

Because meat is an important source of protein for human diets, it is globally produced and processed, while it is projected that this segment of the food industry will expand in the future (*Gómez et al.*, 2019). This will likely result in widening analytical methods for determining organoleptic, physiochemical and food safety parameters required by the market (or by lawmakers that regulate food markets) to ensure consumer acceptance of such meat products. Accordingly, as mentioned earlier, major improvements in laboratory equipment will result in adding additional information to already large datasets from wet chemistry laboratories. These large datasets need to be analyzed in a practical and meaningful way (*Varmuza and Filzmoser*, 2009).

To address this challenge, data analysts employ statistical concepts and tests known as multivariate statistics (*Hidalgo and Goodman*, 2013). The most important aspect of multivariate statistics is the simultaneous testing of multiple independent variables against one (or more) dependent variable(s) to avoid inflating Type I errors. This decreases the corresponding inclination to misleadingly show effects and significances in the dataset that do not actually exist (*Dumancas et al.*, 2015).

In chemometrics, multivariate statistics and data mining are used to draw valid conclusions from large datasets (*Granato et al*., 2018). Multivariate tests include the aforementioned multivariate analysis of variance (MANOVA), numerous factor analyses (e.g., PCA), mathematical modelling, discriminant analysis, etc. (*Dziurkowska and Wesolowski*, 2015). Recently, chemometrics has been used by government agencies and industry to address the challenges of increasingly prevalent food fraud and public concerns about food safety and quality (*Danezis et al.*, 2016). This is in addition to multivariate methods being suitable for determining optimal processing parameters for different production conditions and raw materials (*Granato et al.*, 2018). Accordingly, the application of chemometrics in meat science is only expected to increase as more applications are added to those already mentioned for food safety (*Jurica et al.*, 2021).

7. Conclusion

In conclusion, there are many misconceptions about data analysis in food and meat science, very few of which have been reported in this manuscript. Improper experimental design and data analysis yield data and conclusions of less than optimal quality and diminish the prosperity of the entire field. Most of the methodological principles discussed in this report have existed for a very long time and are well used and known in different scientific disciplines (medicine, epidemiology, etc.). For valid conclusions, it is important to define research hypotheses for a particular test, while data should be collected in the right way and using an experimental design that not only provides useful data, but also saves time and other resources that are very scarce for most researchers around the world. The most frequent types of chemometric tests include MANOVAs, different kinds of factor analysis (e.g., PCA), and mathematical modeling, along with numerous others. When meat scientists and engineers decide to use PCA or mathematical modeling, it should be kept in mind that modeling is used to predict or optimize by some mathematical equation, while PCA is used to find the structure in the data, making it easier for the analyst to deal with large datasets while drawing meaningful conclusions. Both statistical methods have wide applications in the food industry and elsewhere.

Disclosure statement: No potential conflict of interest was reported by the authors.

Funding: This research was funded by the Croatian Science Foundation through the funding of the Hurdle Technology and 3D Printing for Sustainable Fruit Juice Processing and Preservation project, IP-2019-04-2105, and by the Republic of Croatia Ministry of Science.

References

Aldenderfer, M. & Blashfield, R. (1984). *Cluster Analysis*, https://doi.org/10.4135/9781412983648

Ali, Z. & Bhaskar, S. (2016). Basic statistical tools in research and data analysis. *Indian Journal of Anaesthesia*, 60(9), https://doi.org/10.4103/0019-5049.190623

Croarkin, C. (2013). *NIST/SEMATECH e-Handbook of Statistical Methods*

Danezis, G. P., Tsagkaris, A. S., Camin, F., Brusic, V. & Georgiou, C. A. (2016). Food authentication: Techniques, trends & emerging approaches. *TrAC Trends in Analytical Chemistry*, 85, 123–132, https://doi.org/10.1016/j.trac.2016.02.026

Dumancas, G. G., Ramasahayam, S., Bello, G., Hughes, J. & Kramer, R. (2015). Chemometric regression techniques as emerging, powerful tools in genetic association studies. *TrAC Trends in Analytical Chemistry*, 74, 79–88, https://doi.org/10.1016/j.trac.2015.05.007

Dziurkowska, E. & Wesolowski, M. (2015). Multivariate statistical analysis as a supplementary tool for interpretation of variations in salivary cortisol level in women with major depressive disorder. *The Scientific World Journal*, 2015, 1–8, https://doi.org/10.1155/2015/987435

Gómez, B., Munekata, P. E. S., Zhu, Z., Barba, F. J., Toldrá, F., Putnik, P., Bursać Kovačević, D. & Lorenzo, J. M. (2019). Challenges and opportunities regarding the use of alternative protein sources: Aquaculture and insects. In, https://doi.org/10.1016/bs.afnr.2019.03.003

Granato, D., Putnik, P., Kovačević, D. B., Santos, J. S., Calado, V. M. d. A., Rocha, R. S., Cruz, A. G., Jarvis, B., Rodionova, O. Y. & Pomerantsev, A. (2018). Trends in chemometrics: food authentication, microbiology, and effects of processing. *Comprehensive Reviews in Food Science and Food Safety*.

Green, J., Willis, K., Hughes, E., Small, R., Welch, N., Gibbs, L. & Daly, J. (2007). Generating best evidence from qualitative research: the role of data analysis. *Australian and New Zealand Journal of Public Health*, 31(6), 545–550, https://doi.org/10.1111/j.1753-6405.2007.00141.x

Hidalgo, B. & Goodman, M. (2013). Multivariate or multivariable regression? *American Journal of Public Health*, 103(1), 39–40, https://doi.org/10.2105/ajph.2012.300897

Jurica, K., Brčić Karačonji, I., Lasić, D., Bursać Kovačević, D. & Putnik, P. (2021). Unauthorized Food Manipulation as a Criminal Offense: Food Authenticity, Legal Frameworks, Analytical Tools and Cases. *Foods*, 10(11), https://doi.org/10.3390/foods10112570

Larson-Hall, J. (2010). 2.1.1 Levels of Measurement of Variables in A Guide to Doing Statistics. In *Second Language Research Using SPSS* (pp. 33–35). Taylor & Francis.

Putnik, P., Roohinejad, S., Greiner, R., Granato, D., Bekhit, A. E.-D. A. & Bursać Kovačević, D. (2017). Prediction and modeling of microbial growth in minimally processed fresh-cut apples packaged in a modified atmosphere: A review. *Food Control*, 80, 411–419, https://doi.org/10.1016/j.foodcont.2017.05.018

Sohail, A., Al-Dalali, S., Wang, J., Xie, J., Shakoor, A., Asimi, S., Shah, H. & Patil, P. (2022). Aroma compounds identified in cooked meat: A review. *Food Research International*, 157, https://doi.org/10.1016/j.foodres.2022.111385

Szymańska, E., Gerretzen, J., Engel, J., Geurts, B., Blanchet, L. & Buydens, L. M. C. (2015). Chemometrics and qualitative analysis have a vibrant relationship. *TrAC Trends in Analytical Chemistry*, 69, 34–51, https://doi.org/10.1016/j.trac.2015.02.015

Šojić, B., Putnik, P., Danilović, B., Teslić, N., Bursać Kovačević, D. & Pavlić, B. (2022). Lipid Extracts Obtained by Supercritical Fluid Extraction and Their Application in Meat Products. *Antioxidants*, 11(4), https://doi.org/10.3390/antiox11040716

Tabachnick, B. G. & Fidell, L. S. (2007). Principal components and factor analysis. In *Using multivariate statistics (5th ed.)* (pp. 607–615). Allyn & Bacon.

Varmuza, K. & Filzmoser, P. (2009). Chemoinformatics-chemometrics — statistics. In *Introduction to multivariate statistical analysis in Chemometrics* (pp. 1–26). CRC Press Taylor and Francis Group.