Monitoring sensory panel performances with the MAMCAP R-package

Caroline Peltier, Pascal Schlich

CSGA, CNRS, INRA, Université Bourgogne Franche-Comté, 21000, Dijon, France
E-mail : *caroline.peltier@dijon.inra.fr*

Abstract

 A sensory profiling dataset is composed of intensities of sensory attributes given by panelists on several products, most of the time with replications. After a sensory profiling task, the panel leader needs to monitor panel performances to validate the accuracy of the data.

The MAMCAP package (Peltier, Brockhoff, Visalli, & Schlich, 2014) synthesizes discrimination, agreement and repeatability at both individual and panel levels into an HTML table using a user-friendly red-light/green-light code. Indicators of performances can be computed according to the CAP system (Control of Assessor Performances) proposed by Schlich (1997) and popularized by the SensoBase project (www.sensobase.fr).

Optionally, the package proposes to correct the panel heterogeneity in the spread of scoring, the so-called "scaling effect" from the Mixed Assessor Model (Brockhoff, Schlich, & Skovgaard, 2015). In this case, the significance of the scaling effect is tested at both panel and panelist levels and the direction of significant individual scaling effect is indicated in the table.

This paper presents how to use the MAMCAP package with or without taking the scaling effect into account and how to interpret the outputs obtained from the analysis of one dataset taken as an example.

Keywords: MAM, sensory panel performances, scaling

1.   Introduction

Sensory profiling data features a panel of trained judges having scored the intensities of a number of sensory attributes, possibly belonging to several sensory modalities, on a number of products to be compared. The panel leader needs to monitor panelist performances, including at least repeatability, discrimination and agreement with the panel.

Schlich (1997) proposed a synthetic table to represent both panel and panelist performances: the **CAPTable**, in order to summarize all the performance tests in one user-friendly interface. This table was popularized by the SensoBase project, and was consequently used by a large number of panel leaders.

P. M. Brockhoff and Skovgaard (1994) introduced a new performance representing the extent to which one panelist spreads his scores on the scale: the scaling effect. This performance was included in a statistical model: the Mixed Assessor Model (**MAM**) (P. B. Brockhoff, Schlich, & Skovgaard, 2015).

Consequently, Peltier, Brockhoff, Visalli, & Schlich (2014) proposed new performance tests including scaling for panel and panelist. They presented the results of these tests in a table with an interface similar to the CAPTable: **the MAM-CAP Table** and offers an **R-package MAMCAP** allowing any panel leader to obtain his own MAM-CAPTable.

The present paper aims to present how to use this package and interpret its outputs.

2. Material and methods

The R package is freely available on timesens.com (after free inscription). The package must be loaded as a zip file in the R Commander. Then, your data (yourData) must be in the following format: column names as "product", "subject", "rep","attribute 1", …,"attribute p". The order of the columns is not important.
Furthermore, the dataset has to be fully balanced (each subject scored all products the same number of times, and at least once) as a consequence of the use of the Mixed Assessor Model.
To obtain the MAM-CAP Table, the user must submit:

<div align="center">CAPTable(data=yourData)</div>

Then, the MAMCAP table is built as html file in the working directory of R (available with the command getwd()). This table can be copied and pasted in Word file or Powerpoint for any modification of labels.

The lines of MAMCAP are the attributes whereas the columns are the performances and the subjects. They are split in two parts: on the left, the columns Mean, F-Prod, F-Scal, F-Disag and RMSE represent the panel performances; on the right, each column represents one subject.

For the panel performances:
- **Mean** contains the mean obtained for each attribute
- **F-Prod** is the F-statistic of product effect. The cell is green when significant ($p<0.05$), red if not
- **F-Scal** is the F-statistic of scaling effect. The cell is red when significant (heterogeneity of scaling in the panel, $p<0.05$), green if not
- **F-Disag** is the F-statistic of disagreement effect. If the related test is significant (presence of disagreement in the panel, $p<0.05$) the cell is red, green if not
- **RMSE** contains the Root Mean Square of Error of the model and gives indication about repeatability of the panel (No statistic test)

For the subject performances:

- The **color** of the cell gives information about discrimination and agreement. Thus, a **green cell** indicates a discriminative subject in agreement with the panel; **a red cell** indicates a discriminative subject but in disagreement with the panel; and a **yellow cell** indicates a non-discriminative subject (p>0.05).
- < > indicates that the subject used a larger part of the scale than the panel; >< a smaller one; | | else
- The **presence of the letters "lr"** in the cell indicates that the panelist was found less repeatable than the panelist.
- The **dashes** in the cell gives the strength of the discrimination ("---" for p<0.01; " - -" for p<0.05, "-" for p <0.1 and " " if p >0.1)

The attributes are sorted by their discrimination (F-Prod), and the subjects are sorted by their F-Rank representing the discrimination of a subject compared to the others (for one descriptor, the subjects are ordered by their discrimination, and then the average of these rank is calculated). A small F-Rank indicates a more discriminative panelist.

3. Results and interpretation

The outputs are illustrated with an example dataset with 10 panelists, 6 descriptors (on a 0-100 scale), 16 products (coffees) and 3 replicates, complete and balanced (Table 1).

	Mean	F-Prod	F-Scal	F-Disag	RMSE	S8	S2	S10	S3	S4	S7	S1	S5	S9	S6
BURN (Flavor)	46.13	29.12	4.65	0.79	16.84	|---|	<--->	|---|	|---|	|---|	<--->	> lr<	> <	<---lr>	| lr|
BURN (Odor)	42.36	21.79	6.65	0.89	16.57	>---<	<--->	|---|	|---|	<--->	>-<	| lr|	|-|	<--->	| lr|
BITTER (Taste)	44.27	20.02	5.19	0.87	16.72	>---<	<---lr>	|---|	>--<	|---|	<--->	> lr<	|---|	|--lr|	| lr|
ACIDITY (Taste)	32.41	8.47	4.19	1.29	17.19	|---|	<---lr>	> <	|---|	<--->	| lr|	|--|	|--|	| |	| lr|
ACIDITY (Odor)	24.5	5.44	2.39	1.16	16.32	| |	<-->	>-<	|-|	|--|	| |	|---|	<-->	| lr|	< lr>
WOOD (Odor)	13.73	0.59	0.82	1.86	8.4	|---|	| lr|	| |	|---|	| |	|---lr|	| |	| |	| lr|	| |
F Rank	-	-	-	-	-	2.83	3.67	3.83	4.83	4.83	5.83	6.33	6.33	7	9.5

Table 1: MAM-CAP Table (panel limit= 0.05, individual limit=0.05)

Thus, the panel was found discriminative for the attributes except WOOD (**F-Prod column**). A heterogeneity of scaling (**F-Scal column**) was observed for all the attributes except WOOD, and the panel was in disagreement for ACID (Odor) and WOOD (**F-Disag column**). For the descriptor BURN (Flavor), S1, S5 and S6 were found non discriminative whereas the other subjects were found discriminative and in agreement. For WOOD, S8, S2 and S7 were found discriminative but were in disagreement. The presence of "lr" in some cells (BURN Flavor for S1 as an example) indicates that the panelist was found less repeatable than the panel for this attribute.

To conclude, the panel was quite performant on this study, except for the WOOD attribute. S6 seem to need more training. He/she was not discriminative and was less repeatable than the panel.

All the results of the tests (F and p-values) are available in a list obtained by entering:

Res= CAPTable(data=yourData)

When there are less than 3 products and when the dataset is not balanced, running the MAMCAP table is not possible. In this case, CAPTable(data=yourData,option="CAP") allows to produce the classic CAP table without taking scaling effect into account.

Further, as the disagreement test consists in comparing the disagreement of a panelist to unrepeatability, a very repeatable panelist is likely to show disagreement. To raise this issue, a correlation test between the panel and panelist scores can also be run with the inclusion of the parameter correlationTest="Pearson" or "Kendall" in the CAPTable command.

4. Conclusion

The MAMCAP table extends the CAP table by taking the scaling effect into account. It is a useful tool for monitoring panel and panelist performances and is easily available with the MAMCAP package. The MAMCAP package can also produce the classical CAP table. Consequently, with this package any panel leader can obtain an all-in-one table for monitoring his panel and panelist performances.

References

Brockhoff, P. B., Schlich, P., & Skovgaard, I. (2015). Taking individual scaling differences into account by analyzing profile data with the Mixed Assessor Model. *Food Quality and Preference, 39*(0), 156-166.

Peltier, C., Brockhoff, P. B., Visalli, M., & Schlich, P. (2014). The MAM-CAP table: A new tool for monitoring panel performances. *Food Quality and Preference, 32, Part A*(0), 24-27.

Schlich, P. (1997). CAP : une méthode et un outil de contrôle rapide et synthétique des performances des sujets en évaluation sensorielle descriptive. *5èmes journées Agro-industries et Méthodes Statistiques, Versailles, 3-5 Décembre*

Brockhoff, P. B., Schlich, P., & Skovgaard, I. (2015). Taking individual scaling differences into account by analyzing profile data with the Mixed Assessor Model. *Food Quality and Preference, 39*(0), 156-166.

Brockhoff, P. M., & Skovgaard, I. M. (1994). Modelling individual differences between assessors in sensory evaluations. *Food Quality and Preference, 5*(3), 215-224.