

Tutorial for the SensMixed application

Alexandra Kuznetsova, Per Bruun Brockhoff

1. The **SensMixed** package - an overview

The **SensMixed** package is an **R** package for analysing Sensory and Consumer data in a mixed model framework developed by Alexandra Kuznetsova, Per Bruun Brockhoff and Rune Haubo Bojesen Christensen. The package facilitates, among other things:

- analysis of sensory data in a mixed model framework
- novel tools for correcting scaling effects in sensory data
- analysis of consumer data in a mixed model framework
- post-hoc analysis
- plots
- ready to publish output

The package also provides a graphical user interface (GUI), which is based on **shiny R** package Chang et al. (2015), so that it becomes very user friendly and easy to use for sensory practitioners.

2. Obtaining **SensMixed**

You have to have the latest **R** program installed (at least 3.3-0) on your computer before you think about the **SensMixed** package. You can download and install the latest version of **R** from <http://cran.r-project.org/>. Once you have installed **R** you start using it, but many people (and I encourage that) want to install a suitable GUI (graphical user interface) or IDE (integrated development environment). One suitable choice is the (also free) program **RStudio**, which you can download and install from <http://www.rstudio.com/ide/download/desktop>.

Finally you can install the **SensMixed** package by clicking in the **Install** button and writing **SensMixed** in **Rstudio** or write in the **R** console the following command:

```
install.packages("SensMixed")
```

In order to use the functions from the package you need to attach the package by writing the following in your **R** console:

```
library("SensMixed")
```

If you encounter some installation problems, then you are more than welcome to contact the maintainer of the package Alexandra Kuznetsova (alku@dtu.dk).

3. The graphical user interface

The `SensMixed` package contains a `shiny` application Chang et al. (2015), that provides a graphical user interface for the functions contained in the `SensMixed`. In order to launch the application, one simply needs to run the following line in the **R** console:

```
SensMixedUI()
```

This command launches the application in your default web browser. The application supporting the package was designed with focus on simplicity and usability such that valuable information may be accessed in an easy way. Figure 1 represents the main widget of the GUI. In the *Choose type of analysis* panel to the left you can specify which type of analysis you would like to perform. There are two options: analysis of sensory or consumer data. The left panel consists of three tabs: *Input arguments*, *Modeling controls* and *Analysis controls*. In the *Input arguments* you select the names of the variables, that you would like to analyze - these variables are coming from the data. *Modeling controls* tab stands for a detailed specification of the type of modelling. The *Analysis controls* tab stands for specification of the type of analysis to be performed. These tabs are described in details in the following sections. The tabs at the top right of the widget are: *Data*, *Plot output*, *Table output*, *Step output*, *Post-hoc* and *MAM analysis*. The first one stands for the import of the data and is the one that is selected in the figure, the other five tabs are dedicated for the output of results from the analysis of sensory data. In the following sub sections each tab will be explained in detail. First, however, I will explain how the data is imported into the application.

Analysis of Sensory and Consumer data within a mixed effects model framework

This application is a user-friendly interface for the R-package SensMixed

The screenshot shows the main widget of the GUI for the SensMixed application. It features a top navigation bar with tabs for 'Data', 'Plot output', 'Table output', 'Step output', 'Post-hoc', and 'MAM analysis'. The 'Data' tab is active. On the left, there is a 'Choose type of analysis' section with radio buttons for 'Sensory data' (selected) and 'Consumer data'. Below this are sections for 'Input arguments', 'Modelling controls', and 'Analysis controls'. The 'Select attributes', 'Select assessor', 'Select replications', and 'Select products' sections each have a text input field. A blue 'Run Analysis' button is at the bottom left. On the right, the 'Choose data' section has a dropdown menu set to 'Read CSV file from local drive'. Below it, a 'Choose CSV File from local drive, adjusting parameters if necessary' section includes a 'Choose File' button and the text 'No file chosen'. Further down are checkboxes for 'Header' (checked), and radio buttons for 'Separator' (Semicolon selected), 'Quote' (Double Quote selected), and 'Decimal' (Period selected).

Figure 1: The main widget of the GUI of the SensMixed application

4. Data import

In Figure 1 the screenshot for the data import is displayed. As can be seen the user may choose, which data to use: either to import the data from the local files, or the user may choose two data sets, that are contained in the **SensMixed**: the **TVbo** data coming from the sensory studies and the **ham** data coming from the consumer studies - these two data sets I am going to use in this tutorial. The user is always welcome to play around with them.

Even if the first option for choosing the data says: *Read CSV file from local drive*, it accepts different formats:

- plain files such as .txt, .csv
- Excel files such as .xls, xlsx

The type of format is chosen through the *Separator* box. The details of import can be controlled: for example, whether to include a header in the imported data, or which type of decimal to use and others. These options give the flexibilities to import different data.

For illustrative purposes I will use here the **TVbo** data, which is a sensory data, contained in the **SensMixed** package. The **TVbo** data was produced by the highend HIFI company Bang and Olufsen A/S, Struer, Denmark, and was used for a workshop at the 8th Sensometrics Meeting in Norway in 2008. In this data the main purpose was to assess 12 products, specified by two

features: Picture (factor with 4 levels) and TVset (factor with 3 levels). The products were assessed by 8 assessors in 2 replications for 15 different attributes.

The first and foremost step in every analysis is to get the data that is to be analyzed.

5. Choose data

In the *Choose data* tab I select TVbo. Figure 10 shows the screenshot of the chosen data (here TVbo).

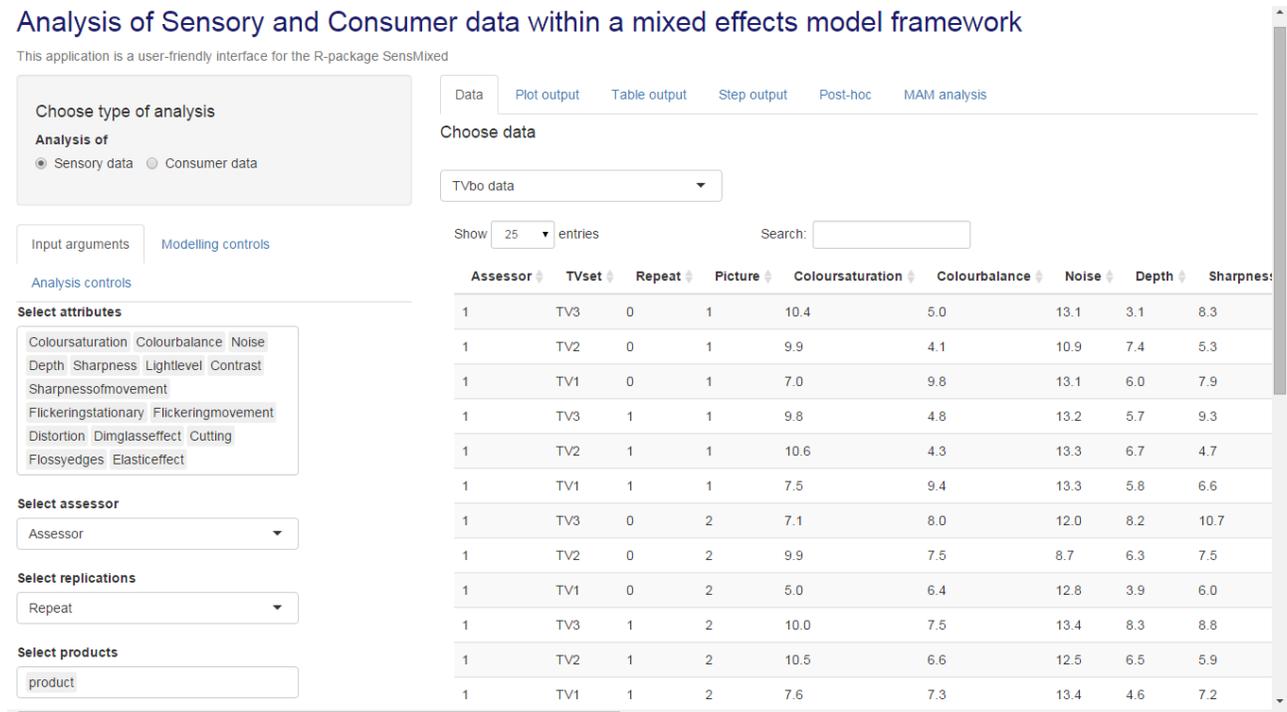


Figure 2: Screenshot for the TVbo data and input arguments for the analysis of the TVbo data

From Figure 10 it can be seen that the TVbo data contains column with the name **Assessor**, which has numbers for each of the eight Assessors. **TVset** column stands for the TVset feature, which has three levels: TV1, TV2 and TV3. **Repeat** column contains two numbers (0 and 1) referring to the number of replication. **Picture** column stands for the Picture feature and contains numbers referring to the levels of the Picture feature: 1,2,3 and 4. **product** column referring to the 12 tested products (3 by 4 TVset Picture combinations). The rest of the columns are the assessors' scores for 15 attributes.

6. Input arguments

In the *Input arguments* tab to the left you need to put the names of the variables for the analysis. For the TVbo data the arguments are automatically filled by the program (see Figure 10):

Select attributes: the names of the columns in the data corresponding to the scores for the attributes. Coloursaturation, Colourbalance, Noise and the rest of 15 attributes

Select Assessor: the name of the column corresponding to the assessors in the data (**Assessor**)

Select replications: the name of the column corresponding to the replication in the data (**Repeat**)

Select products: the names of the product factors. (**product**).

Before clicking the *Run Analysis* button, you may specify what type of analysis you would like to perform via the *Modelling controls* tab.

7. Modelling controls

The analysis of the sensory data in SensMixed is performed in a mixed effects model framework. For each attribute a linear mixed effects model is constructed by using the `lmer` function from the `lme4` package Bates et al. (2013) for constructing the mixed effects models and Kuznetsova et al. (2013) for testing the effects for significance. The *Modelling controls* tab stands for the specification of the type of model to be considered.

Analysis of Sensory and Consumer data within a mixed effects model framework

This application is a user-friendly interface for the R-package SensMixed

Assessor	TVset	Repeat	Picture	Coloursaturation	Colourbalance	Noise	Depth	Sharpness
1	TV3	0	1	10.4	5.0	13.1	3.1	8.3
1	TV2	0	1	9.9	4.1	10.9	7.4	5.3
1	TV1	0	1	7.0	9.8	13.1	6.0	7.9
1	TV3	1	1	9.8	4.8	13.2	5.7	9.3
1	TV2	1	1	10.6	4.3	13.3	6.7	4.7
1	TV1	1	1	7.5	9.4	13.3	5.8	6.6
1	TV3	0	2	7.1	8.0	12.0	8.2	10.7
1	TV2	0	2	9.9	7.5	8.7	6.3	7.5
1	TV1	0	2	5.0	6.4	12.8	3.9	6.0
1	TV3	1	2	10.0	7.5	13.4	8.3	8.8
1	TV2	1	2	10.5	6.6	12.5	6.5	5.9

Figure 3: Screenshot for the Modelling controls for the TVbo data

Figure 3 represents the screenshot for the *Modelling controls* tab. The inputs in this screenshot stand for the specification of a linear mixed effect model for each sensory attribute. In the following I will describe each of the input for the specification of linear mixed effects models in general and will explain later on which inputs I have chosen for the analysis of the TVbo data.

Select product structure: This tab stands for the specification of the fixed effects in linear mixed effects models. There are three following options:

- 1 only main effects form the fixed part

- 2 main effects and 2-way interactions form the fixed part
- 3 all main effects and all possible interactions form the fixed part

Select error structure: This tab stands for the specification of the random effects in linear mixed effects models. There are also three options here:

- ONLY-ASS** assessor effect and all possible interactions between assessor and fixed effects
- ASS-REP** assessor and replicate effect and interaction between them and interaction between them and product effects

Correct for scaling:

- Yes** consider mixed assessor model Brockhoff et al. (2015), where additional fixed effect is added, standing for the scaling effect
- No** do not consider mixed assessor model

8. Analysis controls

Analysis of Sensory and Consumer data within a mixed effects model framework

This application is a user-friendly interface for the R-package SensMixed

Choose type of analysis

Analysis of

Sensory data Consumer data

Input arguments Modelling controls

Analysis controls

Calculate post-hoc

Yes

Simplification of error structure

Yes

Effects to keep in a model

Enter effects separated by space...

Type 1 error for testing random effects

0.1

Type 1 error for testing fixed effects

0.05

Run Analysis

Data Plot output Table output Step output Post-hoc MAM analysis

Choose data

TVbo data

Show 25 entries Search:

Assessor	TVset	Repeat	Picture	Coloursaturation	Colourbalance	Noise	Depth	Sharpness
1	TV3	0	1	10.4	5.0	13.1	3.1	8.3
1	TV2	0	1	9.9	4.1	10.9	7.4	5.3
1	TV1	0	1	7.0	9.8	13.1	6.0	7.9
1	TV3	1	1	9.8	4.8	13.2	5.7	9.3
1	TV2	1	1	10.6	4.3	13.3	6.7	4.7
1	TV1	1	1	7.5	9.4	13.3	5.8	6.6
1	TV3	0	2	7.1	8.0	12.0	8.2	10.7
1	TV2	0	2	9.9	7.5	8.7	6.3	7.5
1	TV1	0	2	5.0	6.4	12.8	3.9	6.0
1	TV3	1	2	10.0	7.5	13.4	8.3	8.8
1	TV2	1	2	10.5	6.6	12.5	6.5	5.9
1	TV1	1	2	7.6	7.3	13.4	4.6	7.2

Figure 4: Screenshot for the Analysis controls for the TVbo data

The input tabs in Figure 4 stand for the specification of the type of analysis to be performed on models defined in input tabs in Figure 3 and are the following ones:

Simplification of error structure: This tab stands for the simplification of the random effects in linear mixed effects models. There are also two options here:

Yes sequentially eliminate non-significant random effects following procedure proposed in Kuznetsova et al. (2015) using the Type 1 error rate (0.1 the default one)

No do not eliminate random effects

Effects to keep in the model: Here one needs to type the effects that one would like to keep in the model even if not being significant. By default *Assessor*) and highest order interaction between **Assessor** and product effects (here **product:Assessor**) are always kept in the model.

Type 1 error rate for testing random effects:

0.1 this option is recommended in Kuznetsova et al. (2015)

0.2

0.05

Type 1 error rate for testing fixed effects:

0.05 this option is recommended in Kuznetsova et al. (2015)

0.01

0.001

9. Analysis of TVbo sensory data. No scaling correction

For illustration purposes I have chosen here *product structure* = 1, which considers only one product main effect (**product**). For the error specification part I have chosen *error structure* = *ASS-REP*, which considers the maximal possible error structure for the initial model, that is all possible random effects and interactions between random and fixed effects. According to the chosen controls for the model specification, the following linear mixed model is constructed for each attribute:

$$y_{ijk} = \mu + a_i + \nu_j + d_{ij} + r_k + ar_{ik} + a\nu_{jk} + \varepsilon_{ijk} \quad (1)$$

$$a_i \sim N(0, \sigma_{assessor}^2), d_{ij} \sim N(0, \sigma_{assessor \times product}^2), r_k \sim N(0, \sigma_{replicate}^2), \\ ar_{ik} \sim N(0, \sigma_{assessor \times replicate}^2), \nu_{jk} \sim N(0, \sigma_{product \times replicate}^2), \varepsilon_{ijk} \sim N(0, \sigma^2)$$

Then the inputs standing analysis controls needs to be chosen. Here, as can be observed from Figure 4, in the *Simplification of error structure* input I have put the default one *Yes*, which eliminates sequentially non-significant random effects as suggested by Kuznetsova et al. (2015). Note that the random effects **Assessor** and **Assessor** \times **Product** are always kept in the model. Finally, I choose the default numbers for the Type 1 error rates.

In order to view the results, I click on the *Plot output* tab (similarly one may open any of the results tabs, but only the first time the analysis is run. Then one needs to click on the *Run analysis* button).

9.1. Plot output



Figure 5: Screenshot for the plot output for random effects for the TVbo data

Figure 5 shows the screenshot of the *Plot output* for analysis of random effects for the TVbo data. In the panel to the left I have chosen the plot for the random effects. Layout multiple means that I get multiple subplots for each random effect, The Plot shows the bars for the sequential χ^2 statistics of the likelihood ratio test applied to each random effect for each attribute. The sequential means that the χ^2 values come from the stepwise selection process of elimination of non-significant random effects. The x-axis stands for the attributes. An overview of the plot indicates that there is no replicate effect in the data. It seems like there is a disagreement between assessors in scoring the products (**product:Assessor** effect is significant for almost all the attributes).

One can tick the tickbox called "stacked", then another version of the plot can be visualized, which can be see in the Figure 6. Here the attributes are ordered by the sum of the chi-square values.

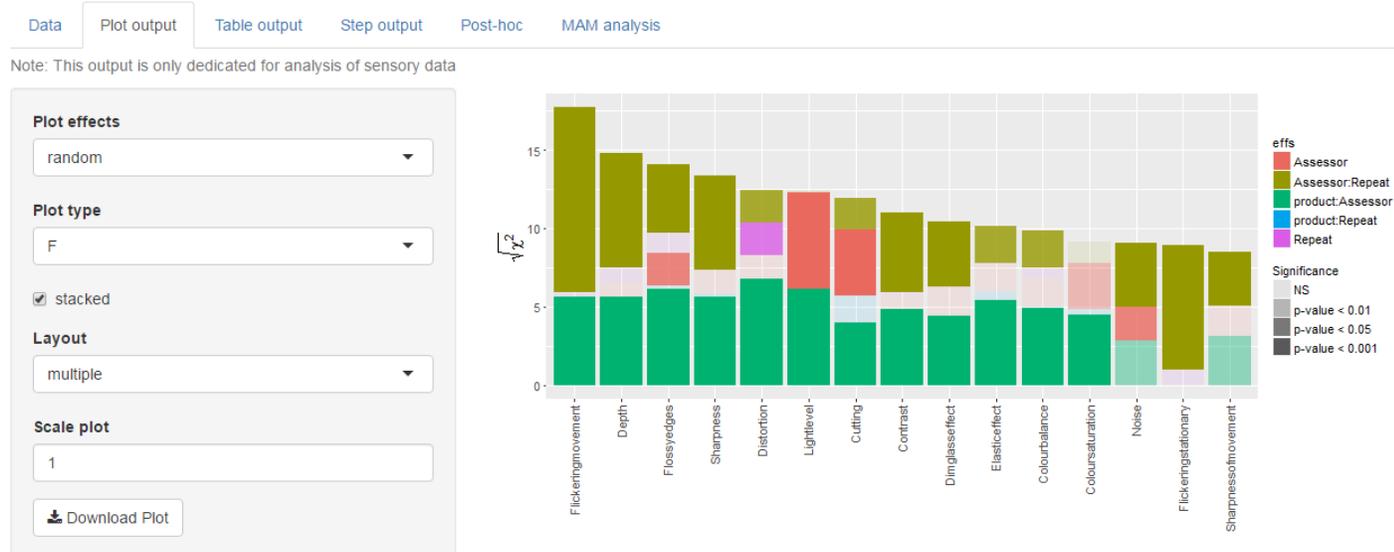


Figure 6: Screenshot for the plot output for random effects for the TVbo data

With the *Download plot* one may easily save the plot in the .pdf format to the local disc. With the *Scale plot* input one may scale the plot - this is valuable when downloading the plot: if it becomes too big, then one may write 2 in the *Scale plot* input and then try to download the plot again.

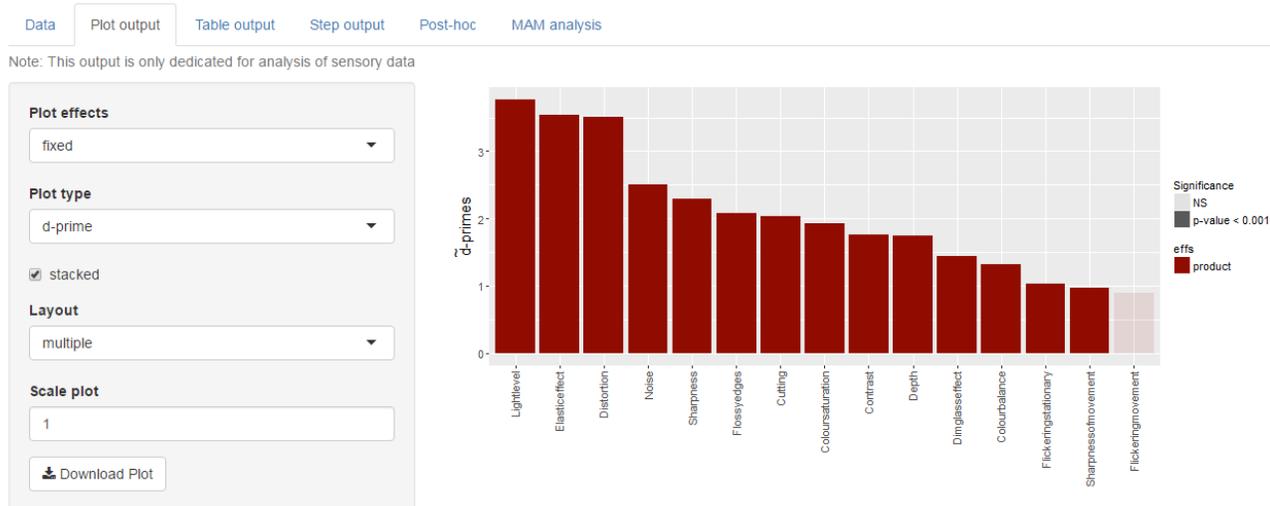


Figure 7: Screenshot for the plot output for the fixed effects for the TVbo data

Then the results from the analysis of the fixed effects can be studied. Figure 7 shows the screenshot for the *Plot output* of the analysis of the fixed effects (**product** effect). The plot visualizes d-tilde for the product effects, the plot is ordered by the size of the d-tilde.

The intensity of the colour corresponds to the p-value for a test of the product effect. From the plot it can be observed, that the `product` effect is significant for all attributes except `Flickeringmovement`. The size of the product effect is the largest for attributes: `Lightlevel`, `Elasticceffect` and `Distortion`.

9.2. Table output

The *Table output* tab, which is next to the *Plot output* tab provides the same output as *Plot output* just in tables.

9.3. Step output

The *Step output* tab provides a detailed information on the analysis of the fixed and random effects. Figure 8 shows the screenshot for the *Step output* tab for the analysis of the `Colourbalance` attribute from the `Tvbo` data. In the panel to the left there is a selection list input: here one may choose for which attribute one wants to view the results of the analysis. Here I have chosen attribute `Colourbalance`. The first table at the top presents the analysis of the random effects. In the `elim.num` column one may view the order in which non-significant effects were eliminated as being non-significant according to the chosen Type 1 error rate (here 0.1). `kept` means the effect was kept in the model. It can be seen that 2 random effects were eliminated: `product:Repeat` and `Repeat`. The next table represents the analysis of the fixed effects. It can be seen that the `product` effect is highly significant, which means that the assessors can discriminate the products according to `Colourbalance` attribute. One can easily save the table to the local disc via *Download Table* button. One may choose to download in *html* format or in *latex* format.

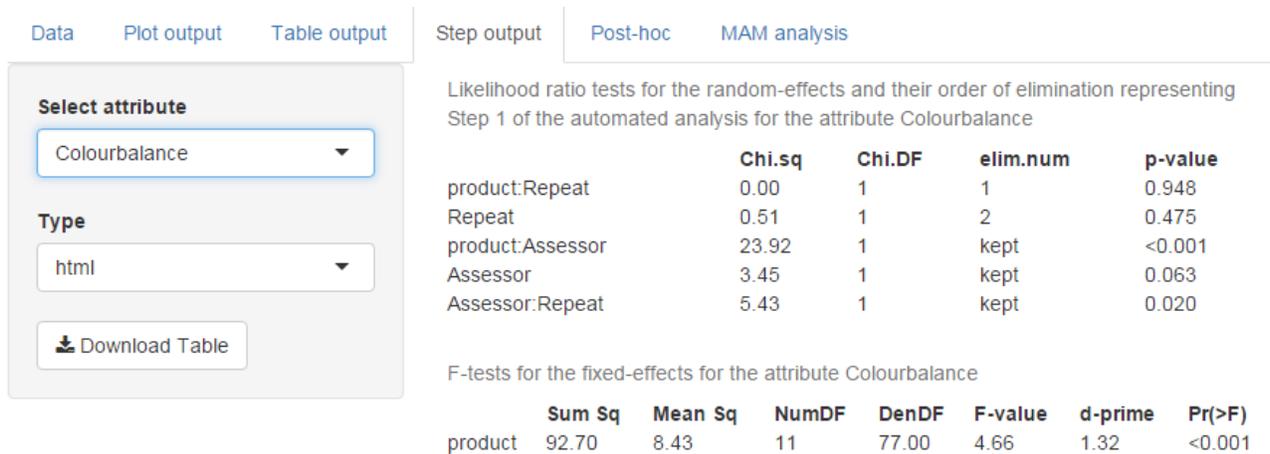


Figure 8: Screenshot for the step output for the `Colourbalance` attribute for the `Tvbo` data

9.4. Post-hoc output

The *Post hoc* tab represents the results of the post-hoc analysis, namely pairwise-comparison tests for the fixed effects. The Figure 9 represents the screenshot for the *Post hoc* tab for the the attribute `Colourbalance`. In the panel to the left there is a selection list input: here one may choose for which attribute one wants to view the results of the analysis. Here I have

chosen the attribute **Colourbalance**. The selection input list *Type of Plot* has a few options: *DIFF of LSMEANS* showing the differences of least squares means for an effect in question and *LSMEANS* showing the least squares means of an effect in question. In selection list *Effects* one can select for which effect to view the results (in this example there is only one **product** effect). The results are displayed in barplots and table, both can be easily downloaded via *Download Table* and *Download Plot* buttons to the local disc. In this example the results are quite hard to interpret since there are 12 products (so $12 \times 11 / 2 = 66$ comparisons are visualized). Considering multi-way product structure, that is instead of only one **product** effect consider two main effects **TVset** and **Picture** and interaction between them **TVset:Picture**, can simplify interpretation and get more insight into the data (Kuznetsova et al., 2015).

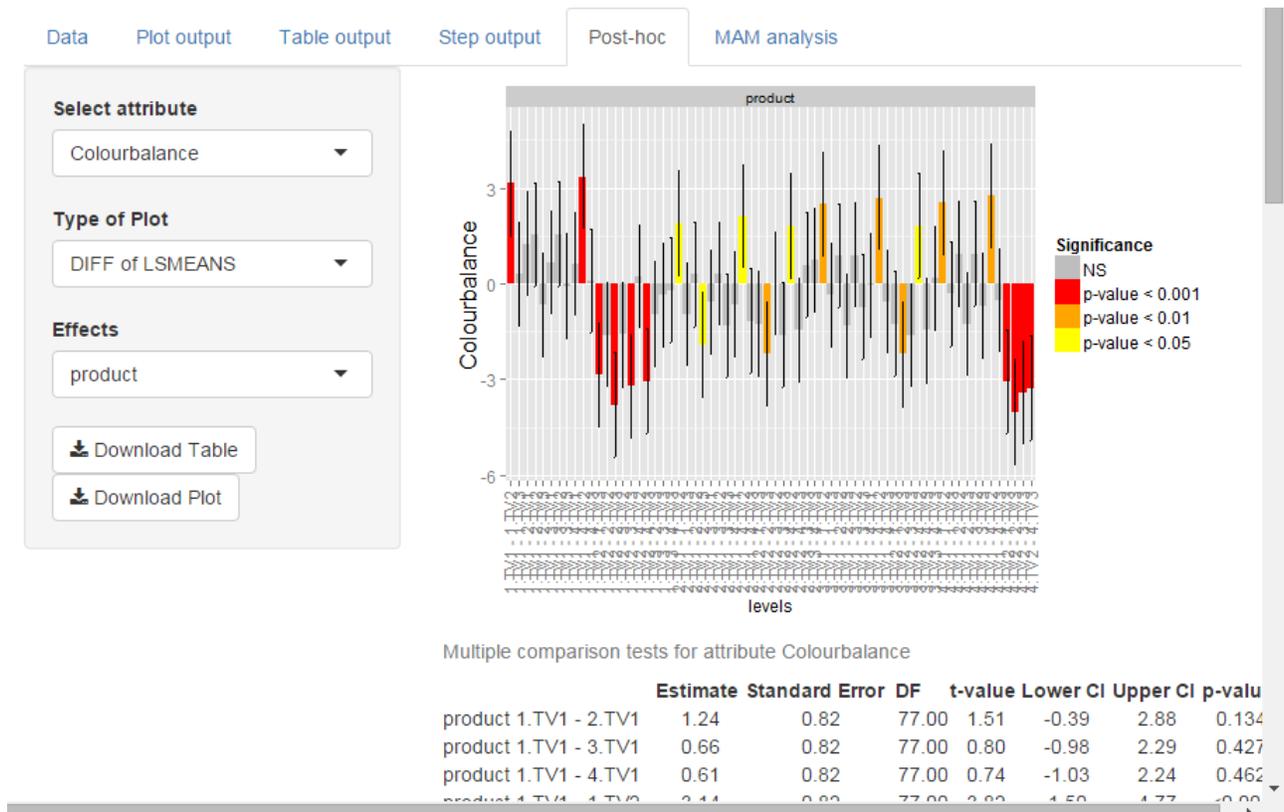


Figure 9: Screenshot for the post-hoc output for the Colourbalance attribute for the TVbo data

10. Analysis of sensory data including scaling correction

In this example the focus is on illustrating how to correct for the scaling effect in the sensory data and how to make the d-tilde plots using the **SensMixed**. For illustrative purposes I will again use here the TVbo data.

In the *Choose data* tab I again select TVbo (see Figure 10).

Analysis of Sensory and Consumer data within a mixed effects model framework

This application is a user-friendly interface for the R-package SensMixed

Choose type of analysis
 Analysis of
 Sensory data Consumer data

Input arguments Modelling controls
 Analysis controls

Select attributes
 Coloursaturation Colourbalance Noise
 Depth Sharpness Lightlevel Contrast
 Sharpnessofmovement
 Flickeringstationary Flickeringmovement
 Distortion Dimglasseeffect Cutting
 Flossyedges Elasticeffect

Select assessor
 Assessor

Select replications
 Repeat

Select products
 product

Data Plot output Table output Step output Post-hoc MAM analysis

Choose data
 TVbo data

Show 25 entries Search:

Assessor	TVset	Repeat	Picture	Coloursaturation	Colourbalance	Noise	Depth	Sharpnes:
1	TV3	0	1	10.4	5.0	13.1	3.1	8.3
1	TV2	0	1	9.9	4.1	10.9	7.4	5.3
1	TV1	0	1	7.0	9.8	13.1	6.0	7.9
1	TV3	1	1	9.8	4.8	13.2	5.7	9.3
1	TV2	1	1	10.6	4.3	13.3	6.7	4.7
1	TV1	1	1	7.5	9.4	13.3	5.8	6.6
1	TV3	0	2	7.1	8.0	12.0	8.2	10.7
1	TV2	0	2	9.9	7.5	8.7	6.3	7.5
1	TV1	0	2	5.0	6.4	12.8	3.9	6.0
1	TV3	1	2	10.0	7.5	13.4	8.3	8.8
1	TV2	1	2	10.5	6.6	12.5	6.5	5.9
1	TV1	1	2	7.6	7.3	13.4	4.6	7.2

Figure 10: Screenshot for the TVbo data and input arguments for the analysis of the TVbo data

The *Input arguments* are then automatically filled, where as product effects one-way factor **product** is chosen. In the *Modelling controls* tab (see Section 7) I select *product structure* = 1, which considers only one product main effect (**product**). Note, that the *MAM analysis* tab view results for only one product effect with at least 3 levels and only for the balanced data. For the error specification part I have chosen *error structure* = *ONLY-ASS*, which considers assessor random effect and interaction between assessor and product effect. In order to correct for the possible scaling effects, I select *Correct for scaling* = *Yes*. For more details about *Input arguments*, *Modelling controls* and *Analysis controls* tabs see Sections 6, 7, 8. According to the chosen controls for the model specification, the following mixed assessor model (MAM) is constructed for each attribute:

$$Y_{ijk} = \mu + a_i + \nu_j + \underbrace{\beta_i x_j}_{\text{scaling}} + \underbrace{d_{ij}}_{\text{disagreement}} + \varepsilon_{ijk} \quad (2)$$

$$a_i \sim N(0, \sigma_{\text{assessor}}^2), d_{ij} \sim N(0, \sigma_{\text{disagreement}}^2), \varepsilon_{ijk} \sim N(0, \sigma^2)$$

where a_i is the assessor main effect, $i = 1, 2, \dots, I$, the ν_j the product main effect, $j = 1, 2, \dots, J$, $x_j = \bar{y}_{.j} - \bar{y}_{...}$ are the centered product averages inserted as a covariate, and hence β_i is the

individual (scaling) slope (the restriction $\sum_{i=1}^I \beta_i = 0$ is imposed in order to ensure that model 2 is uniquely parametrized). The d_{ij} term here captures interactions that are not scale differences hence "disagreements". In (Brockhoff et al., 2015) it was shown that MAM produces valid and improved hypothesis tests for as well overall product differences as post-hoc product difference testing.

Then the inputs standing for analysis controls need to be chosen. Here I have put the default one *Simplification of error structure = Yes*, which eliminates sequentially non-significant random effects. However, the random effects **Assessor** and **Assessor × Product** are always kept in the model, so in this example there will be no elimination of random effects done. Finally, I choose the default numbers for the Type 1 error rates.

In order to view the results, I click on the *Plot output* tab (similarly one may open any of the results tabs, but only the first time the analysis is run. After the second time, one needs to click on the *Run analysis* button whenever the analysis needs to be rerun for the selected inputs).

10.1. Plot output



Figure 11: Screenshot for the plot output for random effects for the TVbo data. Scaling corrected

Figure 11 shows the screenshot of the *Plot output* for analysis of random effects for the TVbo data. In the panel to the left I have chosen the plot for the random effects. The plot shows the bars for the χ^2 statistics of the likelihood ratio test applied to each random effect for each attribute. The x-axis stands for the attributes. An overview of the plot indicates that there is a significant **Assessor** effect for all attributes. Not for all attributes there is a significant **Assessor:Product** interaction. If we compare with the plot in Figure 6, we may notice that the **product:Assessor** effect has become either lower or non-significant for the attributes. This is actually because here the scaling effect is accounted in the models (some of the **product:Assessor** interaction is now taken by the scaling effect, so that **product:Assessor** now represents the pure disagreement between assessors in scoring the products). With the *Download plot* one may easily save the plot in the .png format to the local disc. With the *Scale*

plot input one may scale the plot - this is valuable when downloading the plot: if it becomes too big, then one may write ϱ in the *Scale plot* input and then try to download the plot again.

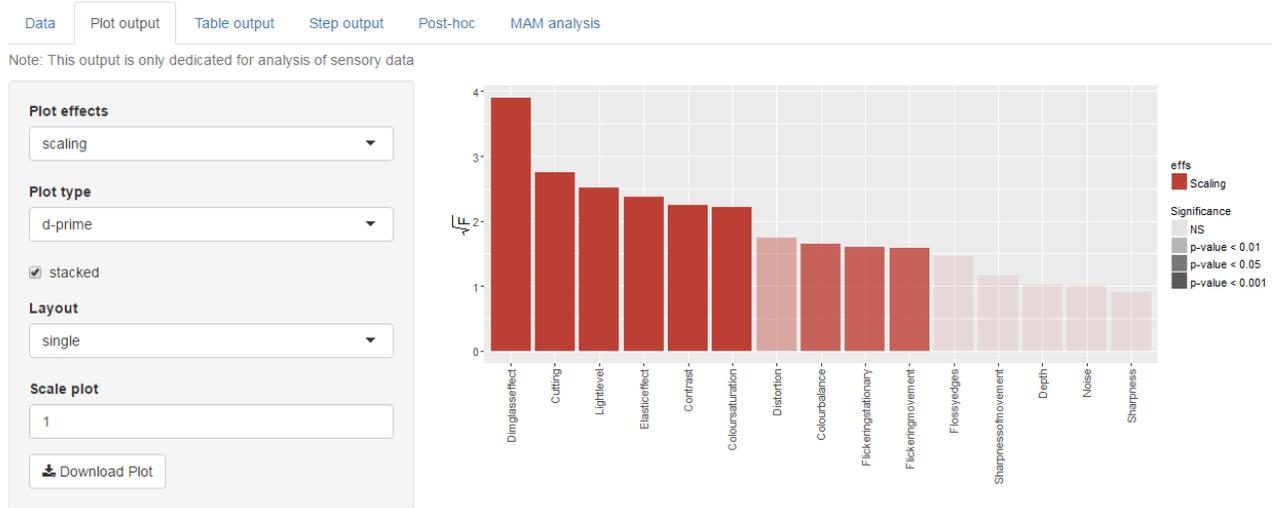


Figure 12: Screenshot for the plot output for the scaling effects for the TVbo data

Next, I take a look at the **Scaling Plot**. The plot shows the bars for the \sqrt{F} statistics of the F test applied to scaling effect for each attribute. The x-axis stands for the attributes. Figure 12 shows the screenshot for the *Plot output* of the analysis of the scaling effects. From the plot it seems like for the majority of attributes the scaling effect is present, hence should be accounted for.

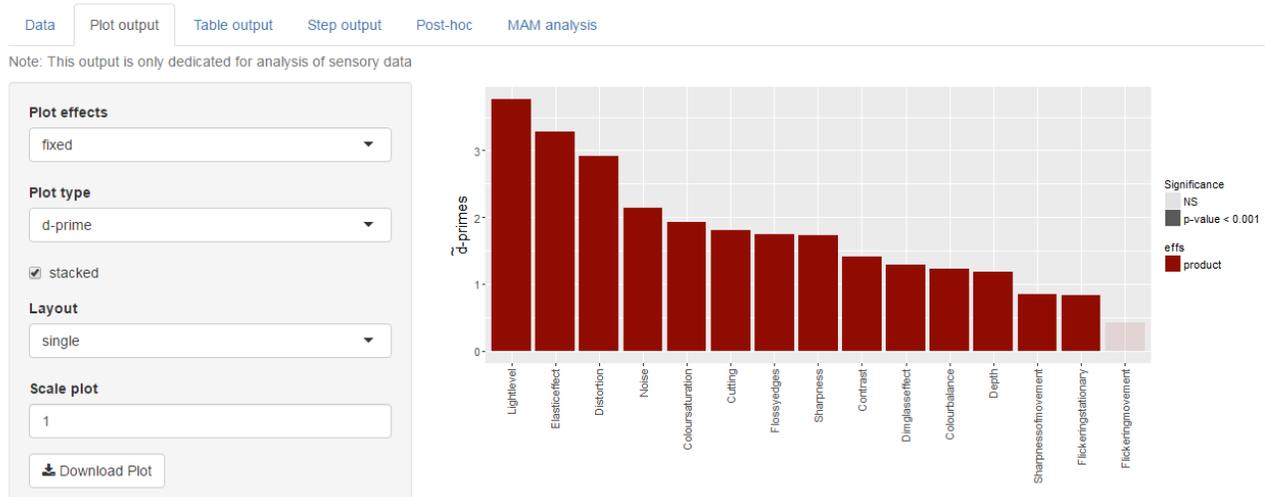


Figure 13: Screenshot for the plot output for the fixed effects for the TVbo data

Finally, I take a look at the analysis of the fixed effect. The plot shows the bars for the \hat{d} of the fixed effect (here product) for each attribute. The x-axis stands for the attributes.

Figure 13 visualizes the screenshot of the \hat{d} in **SensMixed**. From the plot it can be observed, that the **product** effect is significant for all attributes except **Flickeringmovement**. Since \hat{d} represents the effect sizes, the sizes of the bars can be directly compared between the attributes. For example, the size of the **product** effect is the highest for **Lightlevel** attribute. This plot is especially valuable for the multi-way product structure situations.

10.2. Step output

The *Step output* tab provides a detailed information on the analysis of the fixed and random effects. Figure 14 shows the screenshot for the *Step output* tab for the analysis of the **Colourbalance** attribute from the **TVbo** data. In the panel to the left there is a selection list input: here one may choose for which attribute one wants to view the results of the analysis. Here I have chosen attribute **Colourbalance**. The first table at the top presents the analysis of the random effects. In the **elim num** column one may view the order in which non-significant effects were eliminated as being non-significant according to the chosen Type 1 error rate (here 0.1). 0 means the effect was kept in the model. It can be seen that all the random effects are kept in the model. The next table represents the analysis of the fixed effects. It can be seen that both **product** and **Scaling** effects are highly significant. The *p* values, actually, became even lower compared to the model without the **Scaling** effect. One can easily save the table to the local disc via *Download Table* button. One may choose to download in *html* format or in *latex* format.

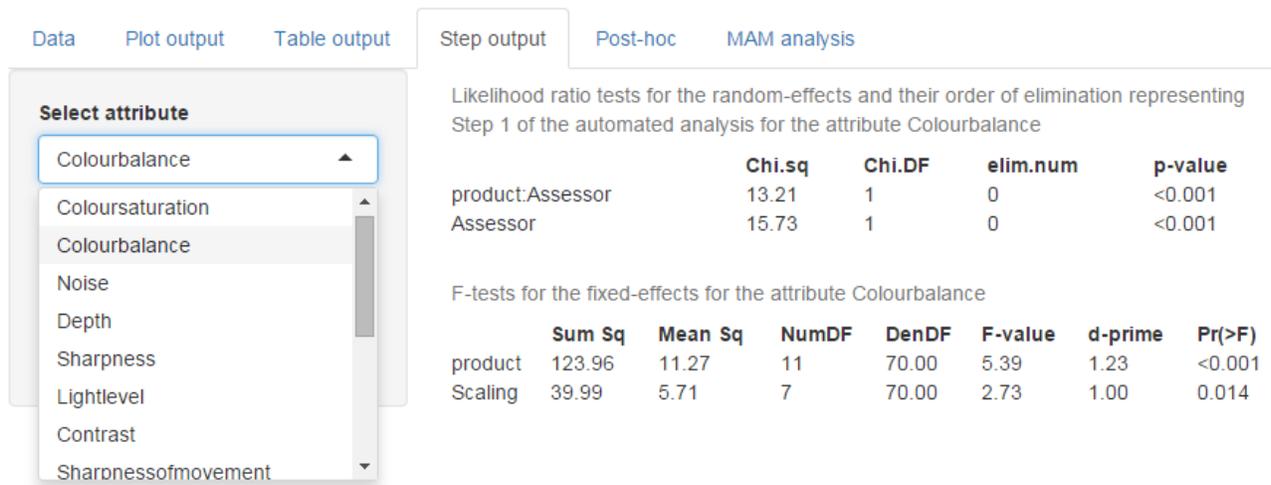


Figure 14: Screenshot for the step output for the **Colourbalance** attribute for the **TVbo** data

10.3. Post-hoc output

The *Post hoc* tab represents the results of the post-hoc analysis, namely pairwise-comparison tests for the fixed effects. The output is similar to the one presented in Figure 9 with no correction of the scaling effect. The difference is only in the standard errors and *p* values in pairwise comparisons. Whenever the scaling effect is significant, the standard errors become lower as well as *p* values, so the tests become more powerful (as also emphasized in Brockhoff et al. (2015))

10.4. MAM analysis output

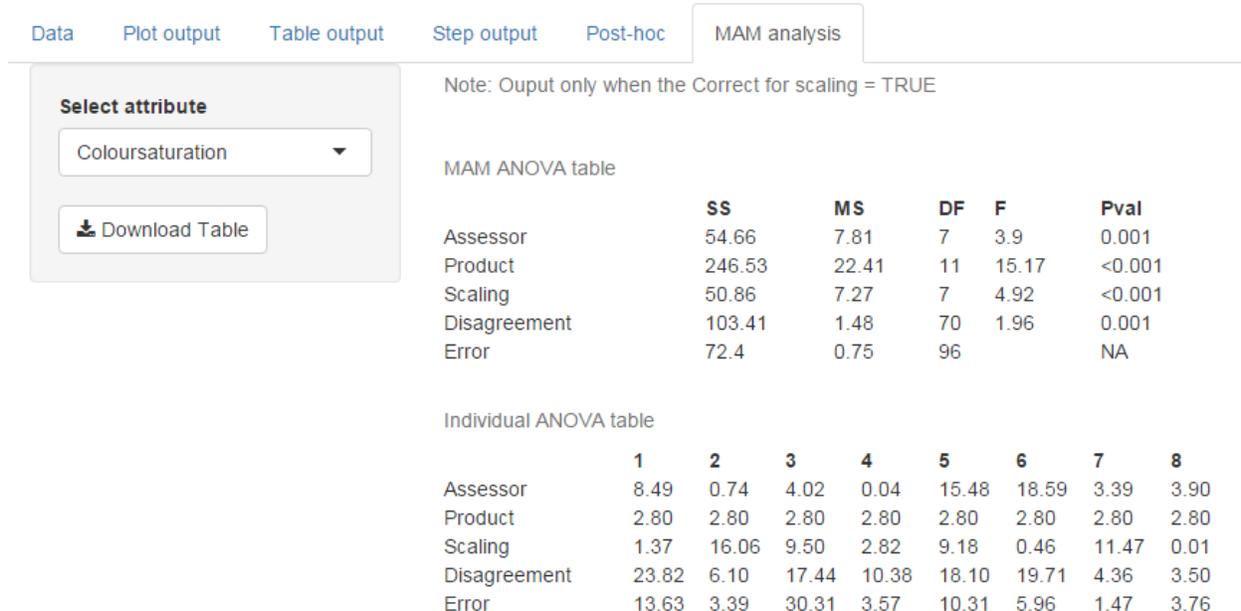


Figure 15: Screenshot for the MAM analysis GUI

In Figure 15 the GUI for the *MAM analysis* tab is visualized. It can be seen that the output is presented for the selected attribute (here *Coloursaturation*). *MAM analysis* tab has some important limitations:

- can only handle balanced data
- can only consider one product factor with at least 3 levels
- can not consider complex error structures (session / batch / carry-over effects e.t.c.)

If one of these requirements is not fulfilled, the the *MAM analysis* tab simply will not produce results. There are a number of table outputs produced for each attribute. In the following these tables will be discussed in details.

10.4.1. MAM ANOVA table

This output is almost the same as the one coming from the *Step output* and gives the overall ANOVA table. One may even check that the F and p values for the scaling and product effects are identical. From Figure 15 it can be observed that both product and scaling effects are significant for the *Coloursaturation* attribute.

10.4.2. Individual ANOVA table

individually decomposed ANOVA table for each attribute, cf. Table 2 in Brockhoff et al. (2015).

10.4.3. Individual performance tests

Figure 16 presents the screenshot for the individual performance tests for the attribute **Coloursaturation** corresponding to Table 2 of Peltier et al. (2014).

Individual performance tests									
	1	2	3	4	5	6	7	8	AVE
Product	1.86 *	1.65 *	1.82 *	4.38 ***	3.12 ***	3.27 ***	2.59 **	3.21 ***	2.74
Scaling	0.79	0.28 ***	1.56 *	1.3	1.55 *	1.12	0.39 ***	1.02	1
Disagreement	1.54	0.78	1.32	1.02 *	1.35	1.4 *	0.66 *	0.59	1.08
Repeatability	1.07	0.53	1.59 ***	0.55	0.93	0.71	0.35	0.56	0.78
Level	0.59 *	-0.18	-0.41	0.04	0.8 **	-0.88 **	-0.38	0.4	0
Correlation	0.67 *	0.53	0.9 ***	0.91 ***	0.9 ***	0.81 **	0.72 **	0.95 ***	0.8

Figure 16: Screenshot for the MAM preference tests for Coloursaturation attribute

P-values are categorized using the usual R-symbols:

- " " = $p\text{-value} \geq 0.1$
- "." = $p\text{-value} < 0.1$
- "*" = $p\text{-value} < 0.05$
- "**" = $p\text{-value} < 0.01$
- "***" = $p\text{-value} < 0.001$

The first four (double) row of these result matrices show the results corresponding exactly to the four rows of Table 2 in Peltier et al. (2014). The MAM-CAP table as such is NOT produced. Instead a descriptive statistic is given for each performance measure. In the following the explanation for each row of the table is provided:

- Product: The square root of the individual assessor product F
- Scaling: The individual beta values (averaging to 1)
- Disagreement: The individual disagreement statistic.
- Repeatability: The individual error (within product) standard deviation
- In addition we provide two more statistics and hypothesis tests:
 - Level: The main effect of assessor (summing to zero) We test whether the individual is different from the average

- Correlation: The correlation between the individual product averages and the overall(consensus) product average. These will average to something close to the so-called Cronbach’s Alpha. We test whether the correlation is different from zero, i.e. it can also be seen as the significance test for negativity (if the correlation is in fact negative)

For more details about the Individual performance tests see Peltier et al. (2014) and the Appendix in Section 13. From Figure 16 it can be seen that the panel is discriminative for the attribute **Coloursaturation** (all the p values for the **Product** effect are less than 0.05). Assessors 2 and 7 seem to use lower scale whereas Assessors 3 and 5 use the upper scale. Regarding the **Disagreement**, assessors 4 5 and 7 seem to disagree with the rest of the panel for the attribute **Coloursaturation**.

10.4.4. MAM based post hoc

Pairwise product differences						
	Estimate	Standard Error	Pval	Lower CI	Upper CI	
1.TV1 - 2.TV1	-0.169	0.43	0.696	-1.0849	0.7138	
1.TV1 - 3.TV1	-0.312	0.43	0.470	-1.2477	0.5603	
1.TV1 - 4.TV1	-1.044	0.43	0.018	-2.1203	-0.1849	
1.TV1 - 1.TV2	-2.681	0.43	<0.001	-4.2288	-1.6521	
1.TV1 - 2.TV2	-2.525	0.43	<0.001	-4.026	-1.5244	
1.TV1 - 3.TV2	-2.006	0.43	<0.001	-3.3509	-1.0818	
1.TV1 - 4.TV2	-3.325	0.43	<0.001	-5.0569	-2.1506	
1.TV1 - 1.TV3	-0.413	0.43	0.340	-1.3626	0.4549	
1.TV1 - 2.TV3	-0.237	0.43	0.582	-1.1624	0.6401	
1.TV1 - 3.TV3	-0.456	0.43	0.292	-1.4134	0.4092	
1.TV1 - 4.TV3	-0.219	0.43	0.612	-1.1412	0.6601	
2.TV1 - 3.TV1	-0.144	0.43	0.739	-1.0568	0.7408	
2.TV1 - 4.TV1	-0.875	0.43	0.046	-1.9127	-0.0179	
2.TV1 - 1.TV2	-2.512	0.43	<0.001	-4.0098	-1.5141	
2.TV1 - 2.TV2	-2.356	0.43	<0.001	-3.8066	-1.3836	
2.TV1 - 3.TV2	-1.838	0.43	<0.001	-3.1316	-0.9316	
2.TV1 - 4.TV2	-3.156	0.43	<0.001	-4.8412	-2.0242	
2.TV1 - 1.TV3	-0.244	0.43	0.572	-1.1694	0.6334	
2.TV1 - 2.TV3	-0.069	0.43	0.873	-0.9733	0.8222	
2.TV1 - 3.TV3	-0.287	0.43	0.506	-1.2191	0.5868	

Figure 17: Screenshot for the MAM post-hoc for Coloursaturation attribute

Figure 17 shows screenshot of MAM based pairwise product comparisons for attribute **Coloursaturation**. The product differences are shown together with the post-hoc p-value. The new method introduced in Brockhoff et al. (2015) is used for calculating confidence limits. The output is also visualized in barplots.

10.4.5. Post-hoc comparison for each product with the mean of the remaining products

Finally, the table is provided for all attributes, where the MAM based post-hoc comparison of each product with the mean of the remaining products for each attribute is performed. Figure 18

Post-hoc comparison of each product with the mean of the remaining products												
	1.TV	12.TV	13.TV	14.TV	11.TV	22.TV	23.TV	24.TV	21.TV	32.TV	33.TV	34.TV
Coloursaturation	-1.22	-1.03	-0.88	-0.08	1.71	1.54	0.97	2.41	-0.77	-0.96	-0.72	-0.98
	***	**	**		***	***	**	***	*	**	*	**
Colourbalance	1.06	-0.29	0.35	0.4	-2.36	-0.61	-0.62	-2.59	0.75	1.79	1.15	0.98
					***			***		**	*	
Noise	2.31	1.49	-2.99	2	-0.12	0.38	-6.52	0.04	1.1	1.53	-1.44	2.22
	***	**	***	***			***		*	**	**	***
Depth	-1.4	-1.7	-0.34	-0.46	-0.53	0.09	-0.05	-1.09	0.25	2.19	1.78	1.27
	**	***						*		***	***	**
Sharpness	-0.62	-1.75	-0.34	0.26	-1.81	-0.44	-1.1	-1.45	1.96	3.07	-0.31	2.54
		***			***		*	**	***	***		***
Lightlevel	-2.71	-3.16	-2.84	-2.1	0.8	0.4	1.64	1.34	1.3	1.41	0.79	3.11
	***	***	***	***	*		***	***	***	***	*	***
Contrast	-0.91	-2.65	-2.19	-0.39	-0.62	0.7	-0.16	0.25	1.62	1.81	0.89	1.65
	*	***	***						***	***	*	***
Sharpnessofmovement	0.23	-0.9	-1.3	1.22	-1.05	0.64	0.08	0.94	0.52	0.19	-2.62	2.04
			*								***	**
Flickeringstationary	0.93	1.77	-1.93	1.26	0.07	0.34	-4.51	-0.35	1.35	1.21	-0.99	0.83
		***	***	*			***		**	*	*	
Flickeringmovement	1.26	0.4	-1.16	0.65	0.23	0.4	-1.39	-0.28	0.87	0.34	-1.81	0.5
	*						*				**	
Distortion	-0.1	-1.58	8.18	0.29	-0.91	-1.35	-0.63	-0.83	-1.26	-0.85	0.24	-1.2
		**	***			**			*			*
Dimglasseffect	-0.87	-1.18	0.8	-1.1	0.18	-1.16	2.2	-0.51	0.4	-0.91	3.1	-0.96
	*	**	*	**		**	***			*	***	**
Cutting	0.96	2.89	1.71	0.85	0.53	0.66	-0.6	1.08	-1.87	-1.96	-2.12	-2.13
	**	***	***	*				**	***	***	***	***
Flossyedges	-1.2	1.56	-1.57	-1.39	-1.37	1.23	-1.57	-1.25	-0.17	5.55	0.45	-0.28
		*	*	*	*		*			***		
Elasticffect	-1.72	-1.61	-0.08	-1.42	-1.27	-0.68	2.46	-0.89	-0.92	-0.64	7.71	-0.94
	***	***		***	**		***	*	*		***	*

Figure 18: Screenshot for the MAM based post-hoc comparison of each product with the mean of the remaining products for each attribute

11. Getting help

To get help on a particular function, e.g. `sensmixed`, you can write `help("sensmixed")` or equivalently `?sensmixed` (or just by typing `sensmixed` in a help tab of RStudio). This works well if you know which function you want help on. And of course you can always use google to search for a particular function.

12. Final remarks

In the analysis of sensory data, quite a lot of random as well as fixed effect can be part of the mixed effects model - this slows down the process of analysis of the data. Depending on the size of the data set, the calculations may take up to few minutes. Whenever one has clicked the *Run analysis* button, in the upper corner a small notification appears, that the calculations

have started and one should wait. If one would like to change the modelling controls of the analysis, then it is important to click the *Run analysis* button in order to run the analysis for the newly selected modelling controls.

References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*. URL: <http://CRAN.R-project.org/package=lme4> r package version 1.0-4.
- 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*, 156–166.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2015). *shiny: Web Application Framework for R*. URL: <http://CRAN.R-project.org/package=shiny> r package version 0.11.1.
- Dahl, D. B. (2014). *xtable: Export tables to LaTeX or HTML*. URL: <http://CRAN.R-project.org/package=xtable> r package version 1.7-4.
- Fai, A. H., & Cornelius, P. L. (1996). Approximate f-tests of multiple degree of freedom hypotheses in generalised least squares analyses of unbalanced split-plot experiments. *Journal of statistical computation and simulation*, *54*, 363.
- Giesbrecht, F., & Burns, J. (1985). Two-stage analysis based on a mixed model: Large-sample asymptotic theory and small-sample simulation results. *BIOMETRICS*, *41*, 477–486.
- Harvey, W. R. (1975). Least-squares analysis of data with unequal subclass numbers.
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2013). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. URL: <http://CRAN.R-project.org/package=lmerTest> r package version 2.0-0.
- Kuznetsova, A., Christensen, R. H., Bavay, C., & Brockhoff, P. B. (2015). Automated mixed {ANOVA} modeling of sensory and consumer data. *Food Quality and Preference*, *40, Part A*, 31 – 38. URL: <http://www.sciencedirect.com/science/article/pii/S0950329314001724>. doi:<http://dx.doi.org/10.1016/j.foodqual.2014.08.004>.
- Kuznetsova, A., de Sousa Amorim, I., & Brockhoff, P. B. (). Analysing sensory data in a mixed effects model 2 framework using the r package sensmixed. *intended for Food Quality and Preference, Food Qual. Preference*, .
- Langsrud, y. (2003). Anova for unbalanced data: Use type ii instead of type iii sums of squares. *Statistics and Computing, Stat. Comput*, *13*, 163–167. doi:10.1023/a:1023260610025.
- Lawless, H. T., & Heymann, H. (2010). *Sensory Evaluation of Food*. Springer Science+Business Media, LLC.
- Næs, T., Brockhoff, P. B., & Tomic, O. (2010a). *Statistics for sensory and consumer science*. John Wiley and Sons Ltd.

- Næs, T., Lengard, V., Blling Johansen, S., & Hersleth, M. (2010b). Alternative methods for combining design variables and consumer preference with information about attitudes and demographics in conjoint analysis. *Food Quality and Preference, Food Qual. Preference*, *21*, 368–378. doi:10.1016/j.foodqual.2009.09.004.
- Nofima Mat, N., Ås (2008). Panelcheck software. URL: www.panelcheck.com.
- 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*, 24–27.
- Satterthwaite, F. (1946). An approximate distribution of estimates of variance components. *BIOMETRICS BULLETIN*, *2*, 110–114.

13. Appendix

Some details on how to do test for each of the individual things based on the individually decomposed ANOVA table. All these tests should be used in an "explorative" manner duly taking into account the multiplicity challenges. Also the tests of individuality of disagreement contribution and repeatability level really lies within models that goes beyond the models expressed above where these effects are assumed homogeneous across assessors.

The idea throughout is that each hypothesis test is to be used together with an observed directly interpretable statistic. Then the single overview table of individual effects will include these interpretable statistics - NOT the hypothesis test statistics, but then the result of performing the hypothesis test is indicated below each interpretable statistics.

13.0.1. Assessor individual product discriminability effect

Finally, the individual product effect is tested by the within-individual product difference F-statistic, as also expressed in the section on ANOVA decompositions above:

$$F_i = \frac{SS_{PROD}^{(i)}/(J-1)}{SS_{Error}^{(i)}/(J(K-1))}$$

and a natural statistic to report here is the square root of the individual F_i :

$$\text{Individual product discriminability statistic: } \sqrt{F_i}$$

13.0.2. Assessor scaling

Test for scaling difference:

$$F_i^{Scal1} = \frac{SS_i^{SCAL}}{SS_{DIS}^{(i)}/(J-2)} \sim F(1, (J-2))$$

which expressed in the fixed scaling model tests:

$$H_0 : \beta_i = 1 \text{ versus } H_1 : \beta_i \neq 1.$$

This is the same test as one would get by doing a simple regression analysis on the individual data averaged over the replicates \bar{y}_{ij} :

$$F_i^{Scal1} = \frac{(\hat{\beta}_i - 1)^2}{SS_{DIS}^{(i)}/(J-2)/(SS_{product}/(KI))}$$

The natural statistic to report here is the individual assessor's scaling value:

$$\text{Individual scale statistic: } \hat{\beta}_i$$

These numbers will average to 1.

13.0.3. Assessor disagreement

Test for disagreement: (also using the individual error rather than the pooled error)

$$H_0 : \sigma_{DIS,i}^2 = 0 \text{ versus } H_1 : \sigma_{DIS,i}^2 > 0.$$

$$F_i^{DIS} = \frac{SS_i^{DIS}/(J-2)}{SS_{Error(i)}/(J-2)} \sim F(J-2, J(K-1))$$

The natural statistic to report here is the individual assessor's disagreement standard deviation:

$$\text{Individual disagreement statistic: } \sqrt{SS_i^{DIS}/(J-2)}$$

This number is comparable with the individual repeatability number given below. The standard deviation here equals the residual standard deviation from a linear regression analysis of the KJ individual scores versus the product average scores. These numbers will average to (approximately) the root of $MS_{Disagreement} \cdot K$.

13.0.4. Assessor repeatability

Test for individual error heterogeneity:

$$H_0 : \sigma_i^2 = \bar{\sigma}_{(i)}^2 \text{ versus } H_1 : \sigma_i^2 \neq \bar{\sigma}_{(i)}^2$$

$$F_i^{ERROR} = \frac{SS_i^{ERROR}}{\frac{1}{I-1} \sum_{\tilde{i} \neq i} SS_{\tilde{i}}^{ERROR}} \sim F(J(K-1), (I-1)J(K-1))$$

The natural statistic to report here is the individual assessor's error standard deviation:

$$\text{Individual repeatability statistic: } \sqrt{SS_i^{Error}/(J(K-1))}$$

13.0.5. Assessor level

Test for Assessor effect:

$$F_i^{Ass} = \frac{SS_i^{ASS}}{MS_{INT}} \sim F(1, (I-1)(J-1))$$

testing (Expressed in a fixed way)

$$H_0 : \alpha_i = \bar{\alpha}$$

Or maybe slightly more relevant:

$$F_i^{Ass} = \frac{I}{I-1} \frac{SS_i^{ASS}}{MS_{INT}} \sim F(1, (I-1)(J-1))$$

testing

$$H_0 : \alpha_i = \bar{\alpha}_{(i)}, \text{ where } \bar{\alpha}_{(i)} = \frac{1}{I-1} \sum_{\tilde{i} \neq i} \alpha_{\tilde{i}}$$

The denominator is the usual interaction mean square:

$$MS_{INT} = \frac{SS_{Scaling} + SS_{Disagreement}}{(I-1)(J-1)}$$

The natural statistic to report here is the individual assessor's average difference to the overall average:

$$\text{Individual level statistic: } \bar{y}_{i..} - \bar{y}_{...}$$

These numbers will average to zero.

13.0.6. Assessor correlation

And hence we can similarly test for the sign of the scaling: (negativity and/or positivity)

$$H_0 : \beta_i = 0 \text{ versus } H_1 : \beta_i \neq 0$$

by the test:

$$F_i^{Scal0} = \frac{\hat{\beta}_i^2}{SS_{DIS}^{(i)}/(J-2)/(SS_{product}/(KI))} \sim F(1, (J-2))$$

These tests are using the individual disagreement variability as error term rather than the pooled disagreement across individuals. The latter would be the natural consequence of the classical variance homogeneity assumptions of the linear mixed models, whereas the former more correctly will account for potential heterogeneities between assessors. The natural statistic to report here (in addition to the individual assessor's scaling value) is the correlation between individual scores and average scores:

Individual agreement statistic: $r_i = \text{c\hat{orr}}((\bar{y}_{\cdot 1}, \dots, \bar{y}_{\cdot J}), (\bar{y}_{i1}, \dots, \bar{y}_{iJ}))$

$$r_i = \text{c\hat{orr}}((\bar{y}_{\cdot 1}, \dots, \bar{y}_{\cdot J}), (\bar{y}_{i1}, \dots, \bar{y}_{iJ})) = \hat{\beta}_i \sqrt{\frac{SS_{Product}/I}{SS_{Prod}^{(i)}}}$$

These numbers will average to something close to the so-called Cronbach's Alpha.