BIONUMERICS Tutorial:

ANOVA and MANOVA

1 Introduction

The central goal of an *analysis of variance* (ANOVA) is to investigate the differences between the means of a (set of) quantitative variable(s) across a number of groups. The main ingredients of an ANOVA are:

- **Explanatory variables:** one or more qualitative variables that determine the group membership of an entry. Therefore, explanatory variables sometimes are called *grouping variables*. An explanatory variable takes values from a finite set of possibilities, i.e. is categorical or binary.
- **Response variables:** one or more quantitative variables. A response variable is treated as a real number.

With only one response variable, the ANOVA is called *univariate*, whereas for more than one response variable the ANOVA is called *multivariate* (or MANOVA).

In this tutorial some of the features of the BIONUMERICS *MANOVA* window will be illustrated using a sample data set (see 2). The different tests and plots present in the *MANOVA* window will not be covered in detail in this tutorial. For detailed information we refer to the BIONUMERICS reference manual.

2 Example data

To illustrate the full possibilities of the *MANOVA* window, a separate data set is made available via the Applied Maths website (https://www.applied-maths.com/download/sample-data, click on "MANOVA sample data").

The sample data set describes an experiment in which the optimal conditions for growth and product formation were determined for a bacterial strain in a broth with a certain carbon source. Two different nitrogen sources were evaluated (yeast extract and ammonium chloride) and three different incubation temperatures (30, 35 and 37°C). These represent the explanatory variables (or grouping variables) in the MANOVA. The experiments were done in 24-well micro titer plates of which four wells were not inoculated. Therefore, 20 replicates are available for each condition. Bacterial growth was evaluated after 24 hours using dry cell weight (in mg/ml) and optical density at 600 nm. The yield of a desired fermentation product was determined using gas chromatography and expressed in mM. The data are available as a tab-delimited text file, designated `MANOVA.TXT`. It is recommended to create a new, separate database.

3 Preparing the database

1. In the *BIONUMERICS Startup* window, create a new database as described in the tutorial ”Creating a new database”.

In the new empty database, we will first create a new character type experiment:

2. In the *Main* window, highlight the *Experiment types* panel and select ***Edit*** > ***Create new object...*** (+).

3. Highlight ***Character type*** and press <***OK***>.

4. Enter a name for the character type, e.g. “Fermentations” and press <***Next***>.

5. Check ***Numerical values***, set two decimal digits and press <***Next***> (see Figure 1).

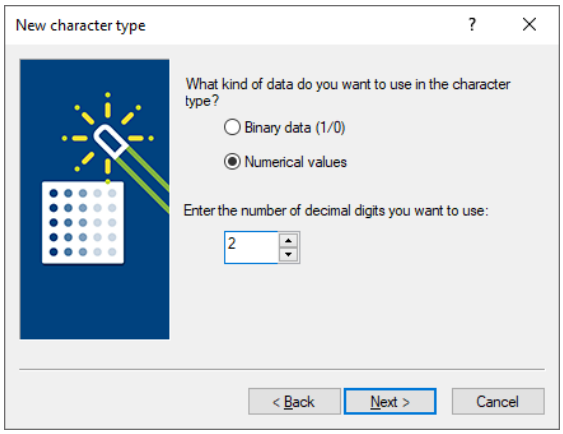

Figure 1: New character type experiment.

6. Set 100 as ***Max value*** and press <***Finish***> to complete the creation of the new character type.

The new character type is added to the *Experiment types* panel.

7. In the *Main* window, select ***File*** > ***Import...*** (, **Ctrl+I**).

8. In the *Import* dialog box, expand ***Character type data***, highlight ***Import fields and characters (text file)*** and press <***Import***>.

9. Browse for the `MANOVA.TXT` file and press <***Next***>.

10. Highlight the row that corresponds to ***Experiment*** and press <***Edit destination***>.

11. In the *Edit data destination* dialog box, highlight ***Key*** and press <***OK***>.

12. Highlight the rows that correspond to ***Temperature***, ***N-source*** and ***Replica*** using the **Shift**-key, and press <***Edit destination***>.

13. In the *Edit data destination* dialog box, highlight ***Entry info field*** and press <***OK***>.

14. In the *Create new* dialog box that appears, leave the default names unaltered and press <***OK***>. Confirm the action.

15. Highlight the three remaining file fields and press <***Edit destination***> again.

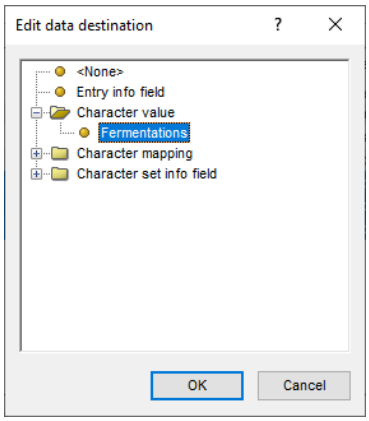

Figure 2: Link to character type experiment.

16. In the *Edit data destination* dialog box, highlight ***Fermentations*** (located under ***Character value***) and press <***OK***> (see Figure 2).

17. In the *Create new* dialog box that appears, leave the default names unaltered and press <***OK***>. Confirm the action.

The *Import rules* dialog box should now look like in Figure 3.

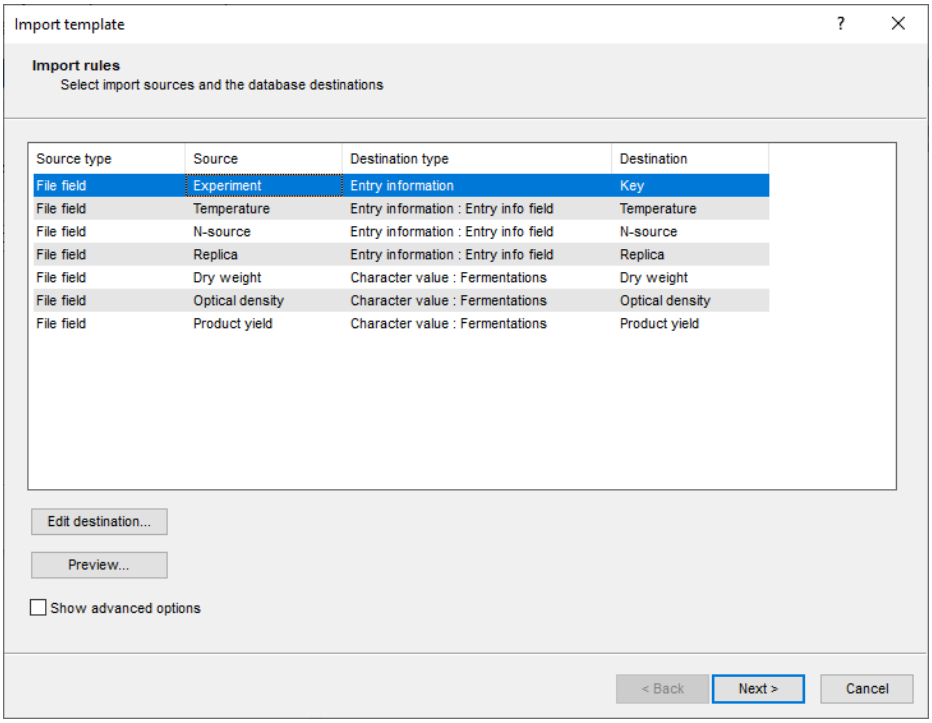

Figure 3: The *Import rules* dialog box after setting up the template for import of the example ”MANOVA.TXT” file.

18. Press <***Next***> and <***Finish***>.

19. Specify a template name (e.g. **Fermentation**) and press <***OK***>.

The template is automatically selected.

20. Press <***Next***> and <***Finish***> to import the information in the database.

The example data are imported in the database (see Figure 4), with the growth conditions (= explanatory variables) as information fields and the growth parameters (= response variables) as character values.

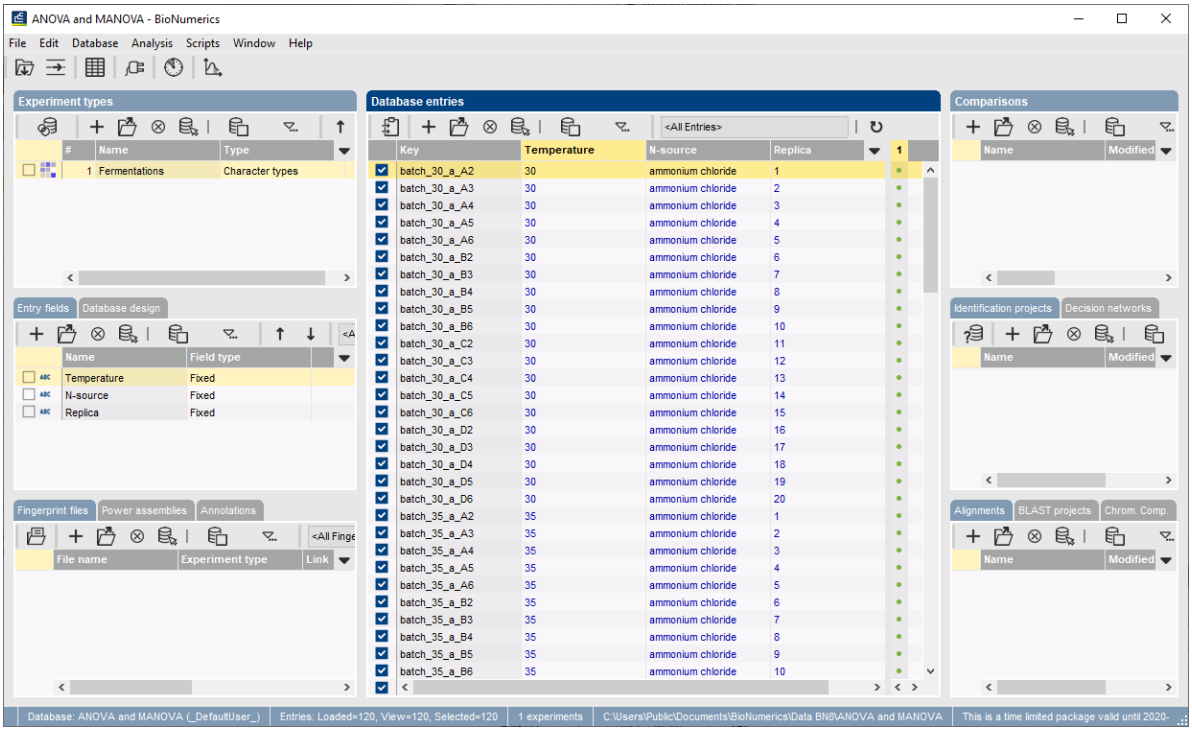

Figure 4: The *Main* window after import of the data.

21. Double-click on the **Fermentations** experiment type in the *Experiment types* panel of the *Main* window.

The *Character type* window opens listing the three characters in the *Characters* panel (see Figure 5).

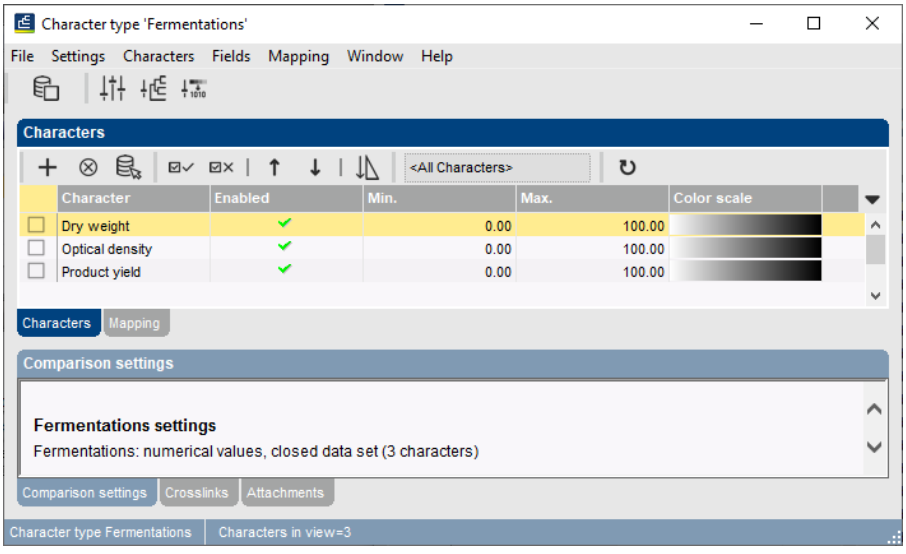

Figure 5: The character type experiment with three characters.

22. Close the *Character type* window.

4 Comparison window

1. In the *Main* window select all entries (e.g. using the **Ctrl+A** keyboard shortcut).

2. Create a new comparison by highlighting the *Comparisons* panel in the *Main* window and selecting ***Edit*** > ***Create new object...*** (+).

3. Select ***File*** > ***Save*** (, **Ctrl+S**). Enter e.g. ***All fermentations*** as comparison name and press <***OK***>.

4. Click on the next to the experiment name **Fermentations** in the *Experiments* panel to display the fermentation data in the *Experiment data* panel.

Initially, the character values are displayed as colors according to the color scale defined for each character.

5. Select ***Characters*** > ***Show values*** ( ) to show the corresponding character values for all entries in the comparison (see Figure 6).

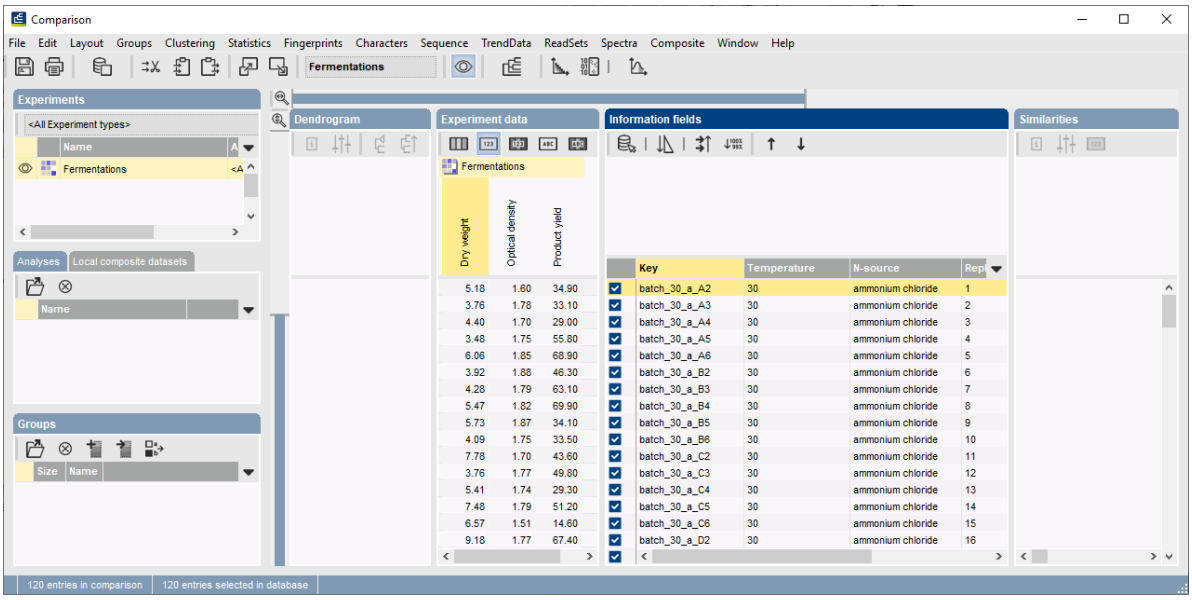

Figure 6: The *Comparison* window.

6. Press the **F4-** key to clear any selection in the *Comparison* window.

5 Performing a MANOVA

1. To start a MANOVA, select ***Statistics*** > ***MANOVA...*** or press the button and select ***MANOVA*** from the menu that appears.

The *Manova analysis* dialog box pops up (Figure 7).

For the example data, we will calculate a *two-way* (with two explanatory variables) *multivariate* analysis of variance (MANOVA):

2. In the *Manova analysis* dialog box, select ***N-source*** and ***Temperature*** as categorical ***Explanatory variables*** and select all characters of the **Fermentations** character type as ***Response variables*** (see Figure 7).

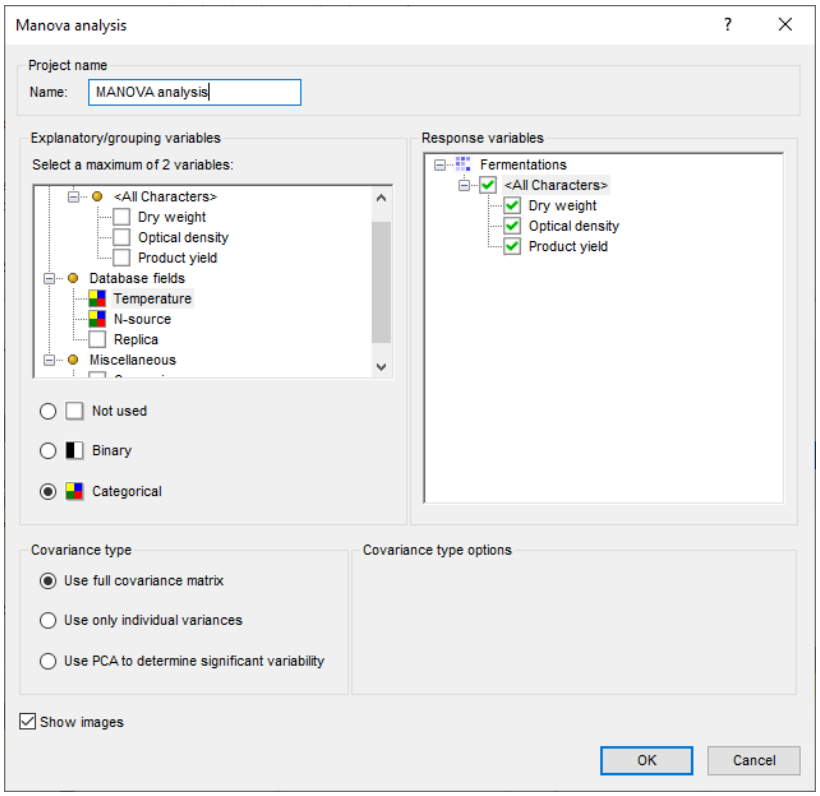

Figure 7: The *Manova analysis* dialog box.

3. Leave ***Use full covariance matrix*** checked and press <***OK***> to calculate a MANOVA analysis with the specified components.

The *MANOVA* window pops up consisting of four different pages, displayed in tabbed view: *Exploratory data analysis*, *Testing model assumptions*, *Analysis of variance*, *Canonical discriminants*.

One can navigate from one page to the other with the ▸ and ◂ buttons (menu commands ***Edit*** > ***Go to next page*** or ***Edit*** > ***Go to previous page***) or by clicking on the corresponding tab.

Each page in the *MANOVA* window contains a number of sections, represented in a hierarchical view. Sections can be collapsed and their content hidden by clicking on the small ”-” (minus) sign that precedes the section name.

5.1 Exploratory data analysis

In the *Exploratory data analysis page* of the *MANOVA* window (Figure 8), the basic elements of a MANOVA are introduced: ***Groups***, ***Group means***, ***Histograms*** and ***Covariance matrices***.

In the groups section of our example data set, it can be seen that the example data set contains in total six groups: three temperature groups (30, 35, and 37°C), multiplied with the two N-source groups (ammonium chloride and yeast extract). Each group contains 20 samples.

4. Click within the groups table (or any other table in the *MANOVA* window). As a result, the table pops up in its own window (see Figure 9).

5. Select ***File*** > ***Exit*** in the *MANOVA Table* window to close it.

For each of the groups in the ANOVA analysis, the mean value for each of the response variables

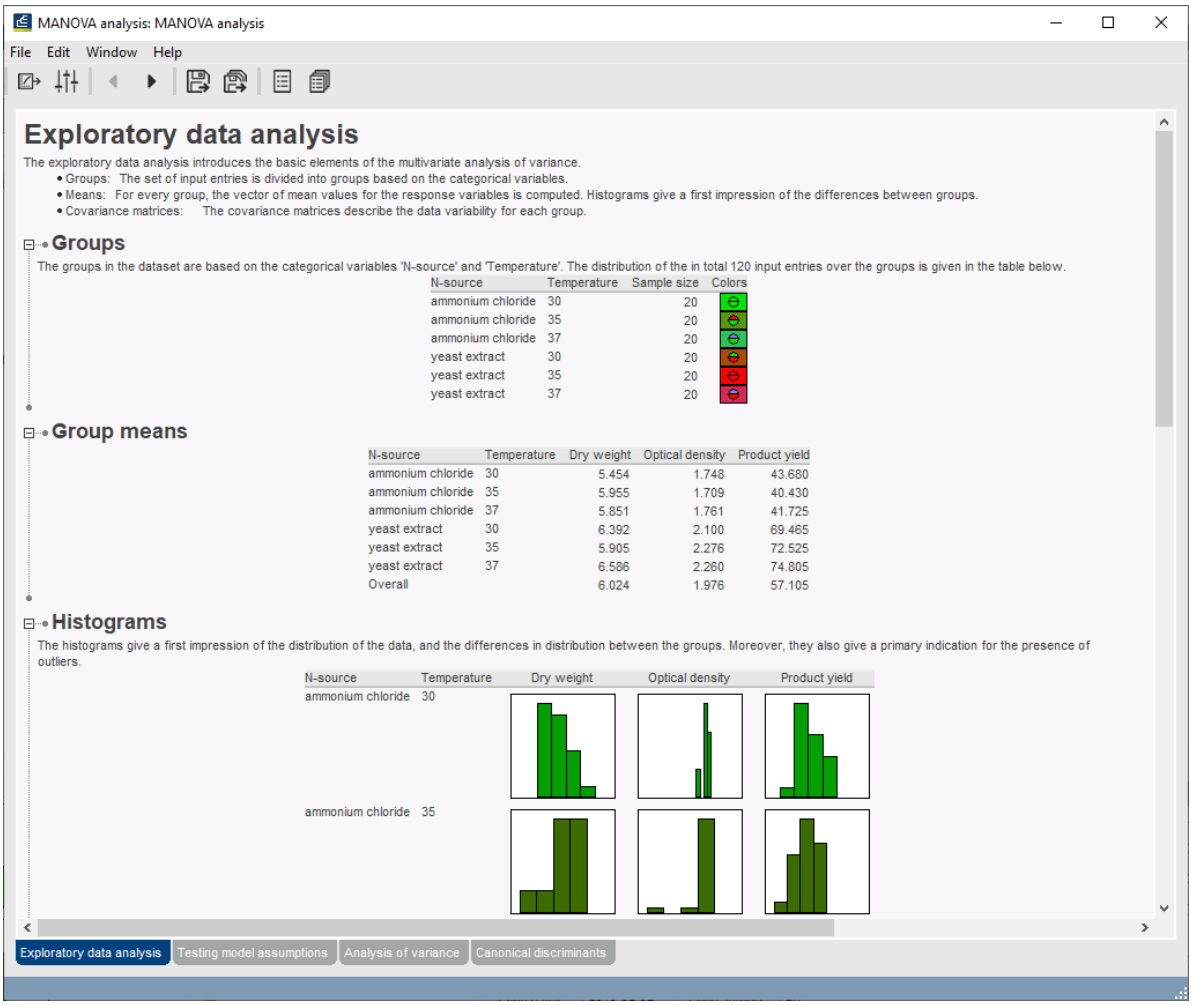

Figure 8: The *MANOVA* window, *Exploratory data analysis page*.

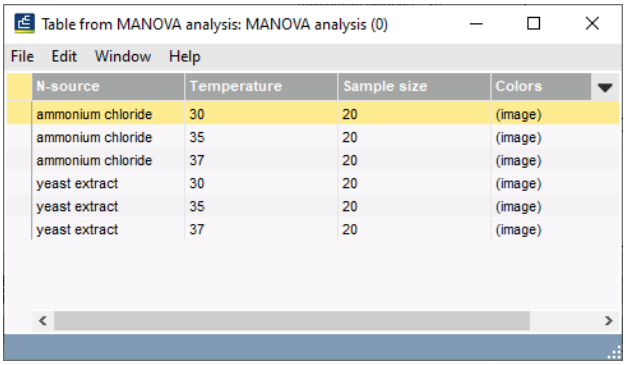

Figure 9: The *MANOVA Table* window, showing the groups in the ANOVA analysis in a grid view.

(in our example data set: Dry weight, Optical density and Product yield) is shown in a table. This gives a first indication how different or similar the group means are.

The group means as such do not give an impression of the spread of the data. Therefore, histograms are drawn for each group - response variable combination. Per response variable, the same X-axis scale is used (global scale). Therefore, the distribution of the values can be visually

compared over the different groups.

6. Click on a histogram, to call the *MANOVA Image* window for the corresponding group.

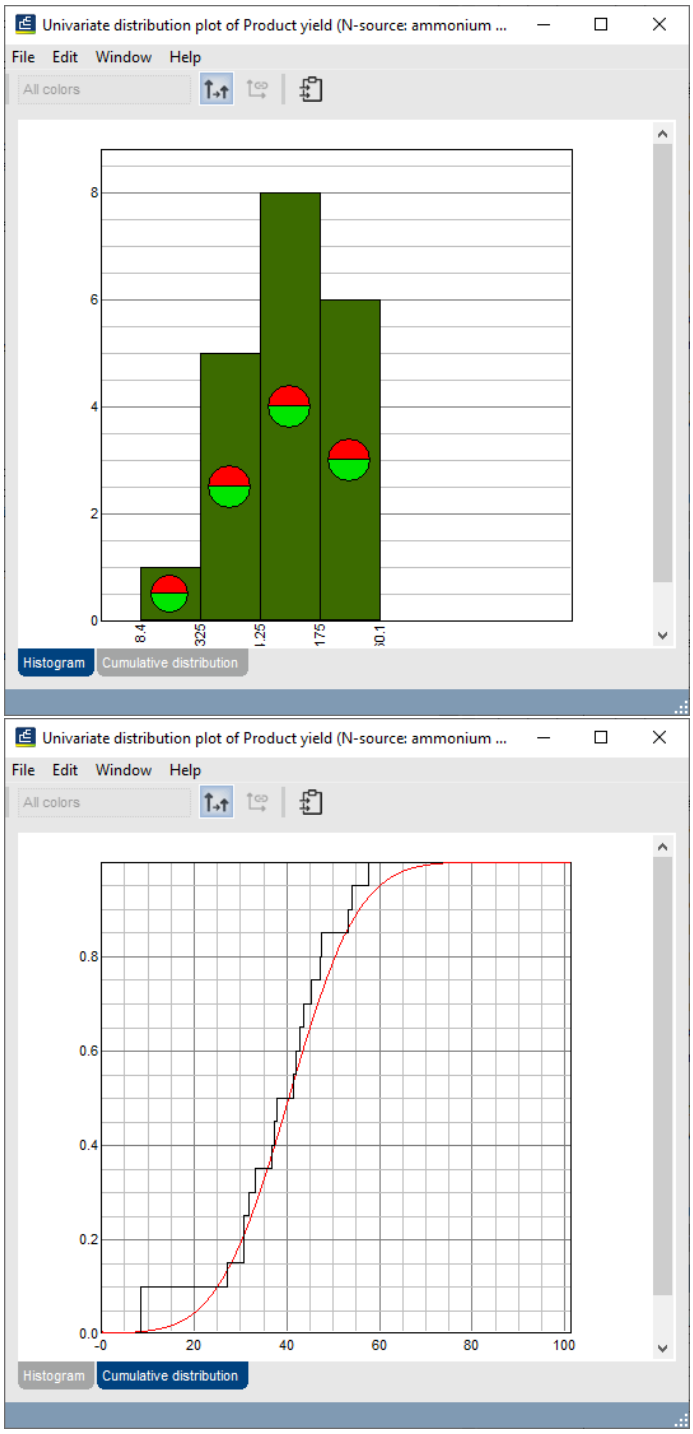

Figure 10: The univariate distribution plot for an ANOVA group in the *MANOVA Image* window, with the *Histogram panel* and the *Cumulative distribution panel* activated.

From the ***Histograms***, we can learn that the within-group variability of the Dry weight measurements is high in relation to the differences between the group means. Therefore, it will be hard to obtain significant conclusions from this variable. Further examination of the histograms reveals a few possible outliers, for example the low value in the Optical density histogram for the 35°C -

ammonium chloride group.

7. Select ***File*** > ***Exit*** to close the *MANOVA Image* window.

For each of the groups in the ANOVA analysis, a matrix of covariances is shown.

8. Click on any of the covariance matrices to pop up the covariance matrix in its own window.

9. Select ***File*** > ***Exit*** to close the window again.

10. In the *MANOVA* window, press the ▸ button to go to the next page: the *Testing model assumptions page*.

5.2 Testing model assumptions

In the *Testing model assumptions page* of the *MANOVA* window (Figure 11), the concordance of the data with the model assumptions is verified. Two assumptions are made: **normality** and **homoscedasticity**.

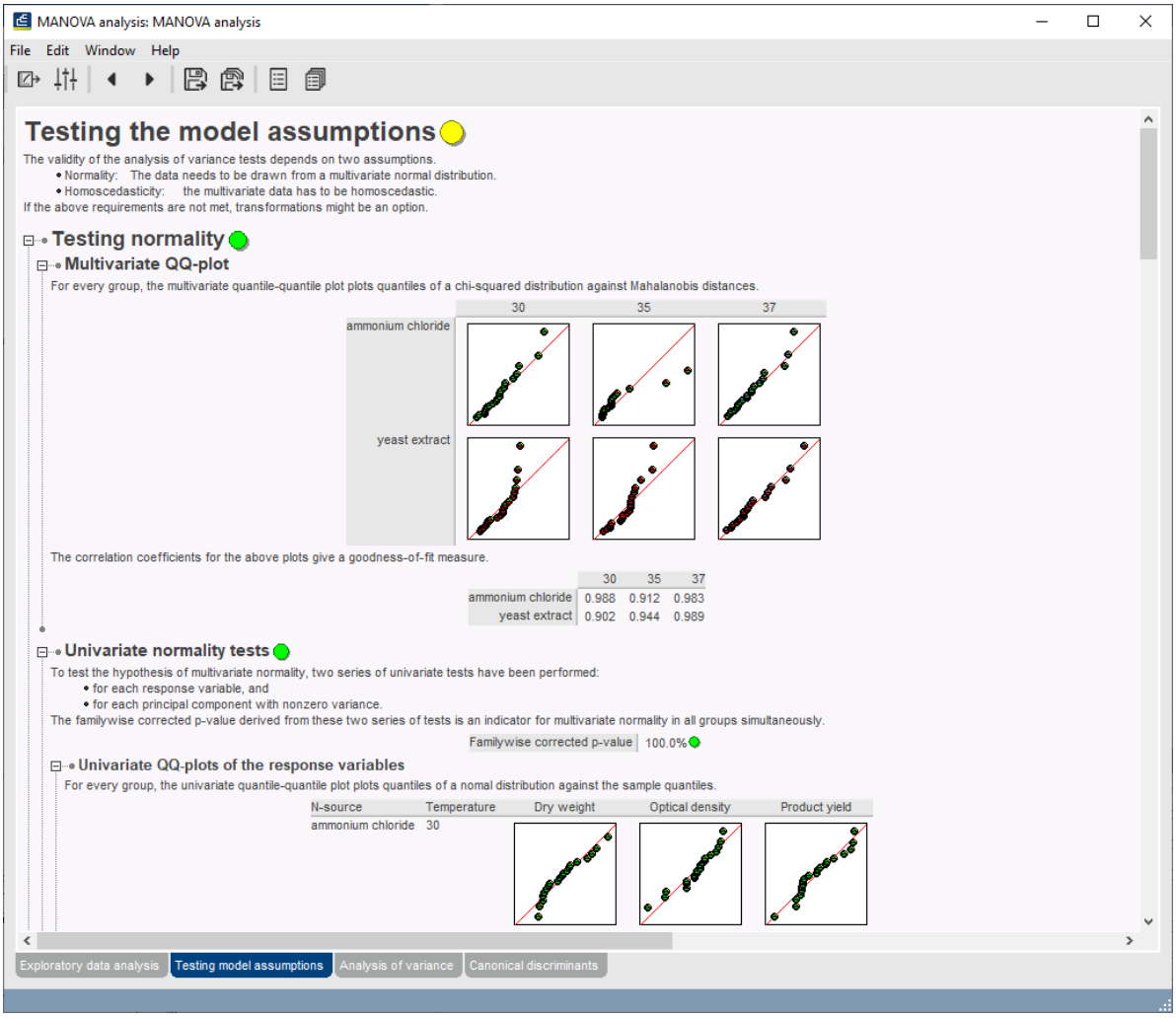

Figure 11: The *MANOVA* window, *Testing model assumptions page*.

For the test of the model assumptions as a whole and for each of the individual tests, a p-value is displayed as a colored dot to the right of the test name. The color of the dot ranges from

green (high probability that the assumption is correct) over yellow to red (low probability that the assumption is correct). When one hovers over the dot with the mouse, the actual p-value is shown.

11. Click on any plot to display it in its own window.

12. Click within any table to open the table in its own window.

In our example data set, outliers can be detected in the 35°C - ammonium chloride group, by looking at e.g. the Univariate QQ-plot of Optical density or at the corresponding correlation coefficient in the table. The same observation can be made in the Univariate quantile plot of Optical density and the p-value of the corresponding KS-test.

Before we can decide to omit or include this outlier, we should have a closer look at the actual data:

13. Open the Univariate quantile plot of Optical density for the 35°C - ammonium chloride group by clicking on this plot in the *Testing model assumptions page*. The plot opens in its own window.

14. In the *MANOVA Image* window, select ***Edit*** > ***Show labels*** > ***Entry key***. The outlier corresponds to the entry with key **batch_35_a_C6**.

15. Go to the underlying *Comparison* window and click on the eye button ( ) next to the **fermentations** character type in the *Experiments* panel to display the character data in the *Experiment data* panel. Show the character values by clicking the button.

By looking at the actual character values for the outlier, it becomes clear that all three values (Dry weight, Optical density, and Product yield) are abnormally low for that specific reaction. Most probably, something went wrong during inoculation of that well, so the reaction can safely be omitted from the analysis.

16. Clear any selection in the *Comparison* window with the **F4** key and select the outlier.

17. Press the button to remove the selected entry from the comparison.

18. Save and close the *Comparison* window.

19. Open the **All fermentations** comparison again and call the MANOVA analysis from the *Analyses* panel.

In the ***Groups*** section on the *Exploratory data analysis page*, it can be seen that the 35°C - ammonium chloride group now contains only 19 samples.

20. Press the button to go to the *Testing model assumptions page*.

It can be seen that the p-values for the tests of the model assumptions are now much better.

21. Press the button to go to the next page: the *Analysis of variance page*.

5.3 Analysis of variance

In the *Analysis of variance page* of the *MANOVA* window (Figure 12), the actual analysis of variance is done.

The following observations can be made for the example data on the *Analysis of variance page* (see Figure 12):

The low p-value in the ***Analysis of variance*** section indicates that the null hypothesis can be rejected: there is at least one pair of groups with a different mean, for at least one of the response variables.

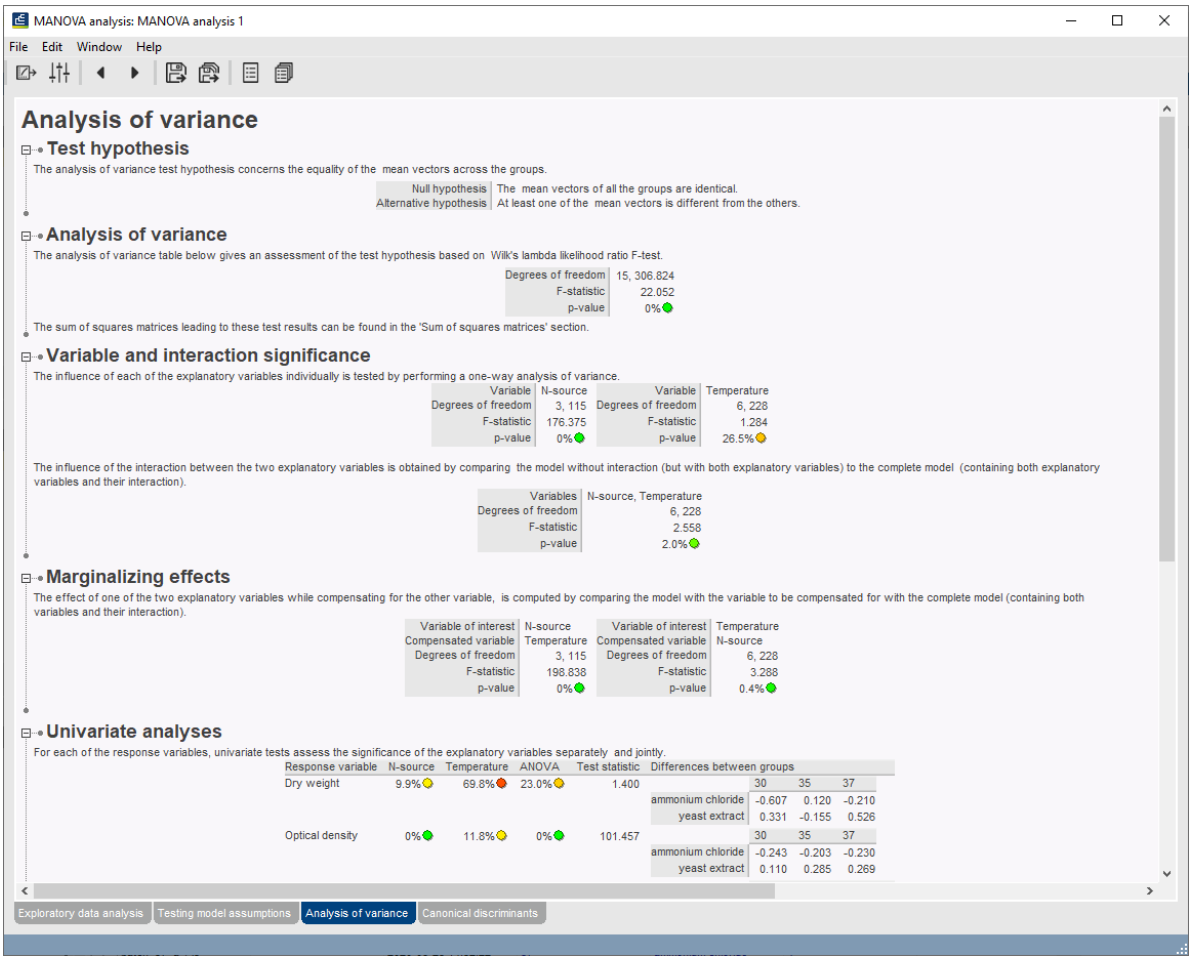

Figure 12: The *MANOVA* window, *Analysis of variance page*.

From the one-way ANOVA done in the ***Variable and interaction significance*** section, it can be concluded that the nitrogen source is more significant than the temperature. Furthermore, both variables are likely to behave independently of each other.

From the ***Univariate analyses***, it can be seen that the N-source is a significant explanatory variable according to the Optical density and Product yield measurements. A possible relation exists between Temperature and Optical density and between N-source and Dry weight. Most probably no relation exists between Temperature and Product yield. Furthermore, it can be concluded that Dry weight does not have much predictive value. This is in concordance with the observation of high within-group variability and the relative small differences between group means made earlier in the *Exploratory data analysis page*.

22. In the *MANOVA* window, press the ▸ button to go to the next page: the *Canonical discriminants page*.

5.4 Canonical discriminants

In the *Canonical discriminants page* of the *MANOVA* window (Figure 13), a canonical discriminant analysis is performed. Discriminant analysis is very similar to PCA, in a sense that it tries to maximize the difference between groups by making linear combinations of the original directions. However, while PCA calculates the best discriminating components without reference to groups, discriminant analysis calculates the best discriminating components for user-defined groups, i.e.

formed by the explanatory variables. The best discriminating components are called discriminants.

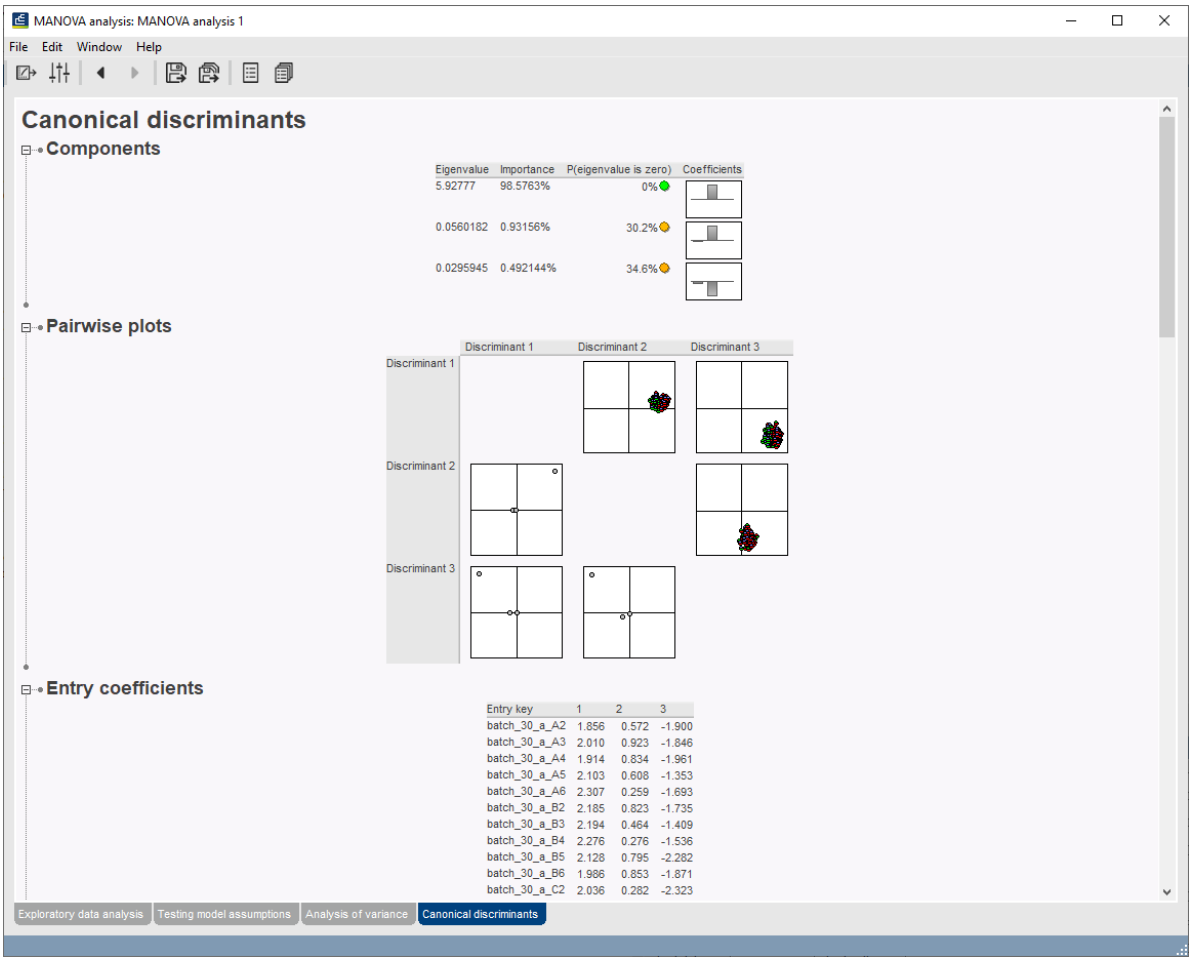

Figure 13: The *MANOVA* window, *Canonical discriminants page*.

Following observations can be made for the example data on the *Canonical discriminants page* (see Figure 13):

In the ***Components*** section, it can be seen that the most important component (or discriminant), i.e. the one that discriminates best between the groups formed by the explanatory variables, relies most on the Optical density response variable.

In the ***Pairwise plots*** where Discriminant 1 is included (1 vs. 2 and 1 vs. 3), a clear separation of the entries according to the N-source is obtained (see Figure 14). In the plot of Discriminant 2 vs. 3, no separation is obtained.

6 Conclusion

The MANOVA analysis performed so far can be the starting point for further (M)ANOVA analyses to explore the example data set, e.g. given a certain N-source (either ammonium chloride or yeast extract), is there an effect of the temperature on the optical density? What happens if the incubation temperatures are categorized differently, e.g. low (30°C) and high (35 and 37°C) temperature? Can we actually learn something from the dry weight or should the dry weight measurements just be omitted from the setup of future experiments?

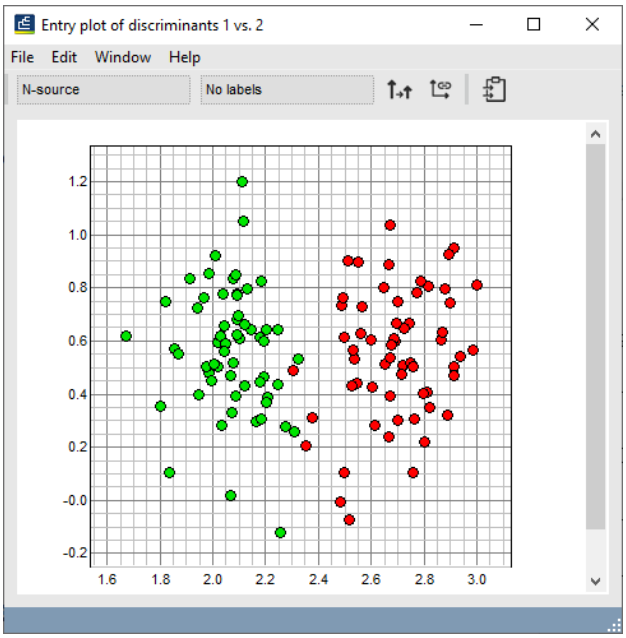

Figure 14: Discriminant 1 plotted versus Discriminant 2 in the example data, colored by N-source.