JMP Instructional data set 034

Sensory Perfume Study

JMP STATISTICAL DISCOVERY

JMP platforms and statistical methods

Graph Builder:	Line graph; dot plot; scatter plot; linear regression
Distribution:	Histogram; summary statistics
Tabulate:	Creating a new data table of summary statistics
Multivariate:	Correlation; Hierarchical clustering
Clustering:	Hierarchical clustering
Fit Y by X:	One-factor ANOVA; multiple comparisons
Fit Model:	Two-factor ANOVA

Problem statement

A manufacturer of cosmetics has conducted a sensory study of their perfume line with the intent of understanding the preferences that different types of consumers have for their products and their aroma characteristics. They currently produce nine different perfumes but may use the results of this study to consider other perfumes to create.

For the first part of the study, their nine perfumes were evaluated by six sensory experts and scored across 11 aroma attributes. The experts evaluated each of the perfumes twice with the 22 sessions performed in a random order with the experts being blind to which perfume they were evaluating. The purpose of this part of the study was to produce a quantitative profile for each of the perfumes.

For the second part of the study 477 consumers who use perfume on a regular basis were recruited. Each of these consumers brought home the nine different perfumes. They were asked to use each perfume on two separate occasions and to do so in a randomly assigned order. They ranked each perfume on a scale of 1-10 on how likely they would be willing to buy the perfume in the future. The consumers were classified into one of 24 different segments based upon their age and the amount of social activity they engage in when they would typically use perfume.

Data

IDS_034_Experts Scores.jmp

Perfume	Names of the nine perfumes
Evaluator	Names of the six expert evaluators
Session	Identification of the two evaluation sessions
Sweet	Score on the presence of sweet aroma notes (1-10)
Fruity	Score on the presence of fruity aroma notes (1-10)
Floral	Score on the presence of floral aroma notes (1-10)
Aquatic	Score on the presence of aquatic aroma notes (1-10)
Green	Score on the presence of green aroma notes (1-10)
Citrus	Score on the presence of citrus aroma notes (1-10)
Aromatic	Score on the presence of aromatic aroma notes (1-10)
Woody	Score on the presence of woody aroma notes (1-10)
Amber	Score on the presence of amber aroma notes (1-10)
Vanilla	Score on the presence of vanilla aroma notes (1-10)
Spicy	Score on the presence of spicy aroma notes (1-10)

IDS_034_Consumers Scores.jmp

ID	Identification of the consumer
Age	Age group of the consumer
Social Activity	Amount of social activity the consumer engages in when they would typically use perfume
Age_SA	Variable combining the 24 different consumer groups based on age and amount of social activity
Amour Wind	Score on willingness to buy Amour Wind perfume (1-10)
Ivory Reign	Score on willingness to buy Ivory Regin perfume (1-10)
Jasmique	Score on willingness to buy Jasmique perfume (1-10)
Lavora	Score on willingness to buy Lavora perfume (1-10)
Muskara	Score on willingness to buy Muskara perfume (1-10)
Nocturne	Score on willingness to buy Nocturne perfume (1-10)
Oceanea	Score on willingness to buy Oceanea perfume (1-10)
Saltflower Veil	Score on willingness to buy Saltflower Veil perfume (1-10)
Zestara	Score on willingness to buy Zestara perfume (1-10)

Exercises

Exercises 1-3 will use the 'IDS_034_Expert Scores' data set.
Exercises 4-6 will use a new data set that you will create in Exercise 3
Exercises 7-13 will use the 'IDS_034_Consumer Scores' data set.
Exercises 15-16 will use a new data set that you will create in Exercise 14

1. Create a set of comparison line graphs displaying the aroma scores for each perfume by all expert evaluators. Do the evaluators have similar scores across the attributes, or is there a significant amount of disagreement? Are there any aroma attributes for any of the perfumes where there is complete agreement across all evaluators? What is the largest difference observed? Identify the perfumes that have the most and least amount of agreement across the evaluators?

   Instructions:
   Graph > Graph Builder. Place all 11 aroma variables on the X axis.
   Place 'Perfume' in the Wrap role and 'Evaluator' in the Overlay role.
   Choose the line graph icon. 

   Create a second graph that directly displays the range in the scores for each aroma attribute and perfume that can help more directly show the level of agreement and will be used for the next exercise.
   Launch Graph > Graph Builder. Place all 11 aroma variables on the X axis.
   Place 'Perfume' in the Wrap role. Choose the scatterplot graph icon. 

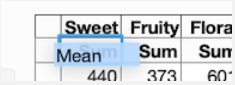
   Choose to display the difference between the largest and smallest score (i.e., range) across all 12 values for each attribute by selecting Range from the Summary Statistic drop down menu in the controls on the left.

2. Create a graph displaying the differences in the two scores made by the evaluators for each attribute for each of the perfumes. Are the two scores similar or are there large differences?

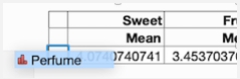
   Instructions:
   From the second graph created in Exercise 1 (ranges), select Local Data Filter under the top red triangle. Choose 'Evaluator' and click the "+" symbol. Individual evaluators can now be selected and the difference in their two scores for each attribute will be displayed.

3. Create a new data table that is the average score across all 6 x 2 = 12 values for each attribute for each perfume.

Instructions:
Analyze > Tabulate. Drag all 11 aroma attribute variables into the Drop zone for columns. Select and drag the Mean from the summary statistic list on top of the word Sum in one of the columns in the table to use that as the summary statistic to display.



Select and drag the variable 'Perfume' and place in the row drop zone.



Change the format to display the data to the 10th decimal place by selecting the Change Format button in the bottom left. Select the "Use the same decimal format" check box. Choose Fixed Dec and 1 for Number of decimals.

Choose Make into Data Table from the top red triangle. Save the file using a descriptive name such as 'Avg Expert Evaluators Aroma Scores'.

Exercises 4-6 will use this new data set created from the summary statistics table

4. Create a line graph that displays the average scores for each attribute and perfume. Based upon examining these aroma profiles, identify a few pairs of perfumes that are most like each other. Note: We will perform a more formal analysis in exercise 6 to group perfumes that are similar across their aroma attributes.

Instructions:
Using the new data set created in Exercise 3.
Graph > Graph Builder. Place all 11 aroma variables on the X axis.
Place 'Perfume' in the Wrap role. Choose the line graph icon. 

An alternative version of the graph, which can help make comparing the perfumes easier, is to use 'Perfume' in the overlay role instead. From this new graph, choose Local Data Filter under the red triangle and select 'Perfume' and hit the "+" symbol. Now an individual perfume or multiple perfumes can be selected and only those will be shown on the graph. It is also helpful to lock the scales. Do so my choosing Lock Scales under the red triangle next to the Graph Builder title.

5. Graphically and numerically display the correlations among all pairs of the 11 attributes (Note: There are 66 pairs). Identify 2-3 pairs of attributes that have a strong positive correlation (> 0.85) and 2-3 pairs that have a strong negative correlation (< -0.85). Provide an interpretation of what it means that these pairs of aroma attributes are highly correlated. Identify all pairs that have very little to no correlation (between -0.20 and +0.20). Which attribute are these most common with? Provide an interpretation of this result.

> ***Instructions:***
> *Analyze > Multivariate Methods > Multivariate. Place all 11 aroma variables in the Y, Columns role. Click OK. To combine the numeric correlation values with the scatterplots, choose Matrix Options > Heat Map, then Show Correlations under the red triangle next to the Scatterplot Matrix title.*

6. Perform a hierarchical cluster analysis putting the 9 perfumes into 5 separate clusters. Describe these clusters via. the scores across the attributes. Is this similar to your conclusions from Exercise 4? If a consumer liked one of the perfumes in a given cluster, what would be your expectation be of their liking another perfume in that cluster? For example, if a consumer liked Nocturne, would you expect them to also like Muskara? Would it necessarily follow that they would most likely dislike a perfume in a different cluster (e.g., Amour Wind)? Why or why not?

> ***Instructions:***
> *Analyze > Clustering > Hierarchical Cluster. Place all 11 aroma variables in the Y, Columns role. Place 'Perfume' in the Label role. Click OK. Choose Color Clusters from the top red triangle. Drag the diamond symbol on the dendogram until 5 clusters are created. Choose Cluster Summary and Parallel Coord Plot under the top red triangle. Note: The Parallel Coordinate Plots are scaled so that the lines are drawn relative to the values for the other perfumes for a particular aroma. For example, note that for the Zestara perfume, the values for Fruity and Citrus are close to the top of graph. The values for these variables are quite different (4.6 for fruity and 8.0 for citrus). These, however, are the largest values for those two aromas as compared to all perfumes. Use the graphs created in Exercise 4 to see the actual values on the 1-10 scale.*

Exercises 7-13 will use the 'IDS_034_Consumer Scores' data set.

7. Summarize the 'willingness to buy' scores from all 477 consumers for each the 9 perfumes. Do any of the perfumes score consistently high or low or is there a large amount of variation with some consumers very likely to use a given perfume while others very unlikely to use?

> ***Instructions:***
> *Analyze > Distribution. Place all 9 perfume variables in the Y, Columns role. Click OK.*

8. Generate a table displaying the correlation among all pairs of the 9 perfumes (Note: There are 45 pairs). Create a scatter plot visualization that will allow you to display the scores from any pair of perfumes. Examine the correlation values as well as the scatterplot between perfumes that are in similar clusters as determined in Exercise 6. Is this aligned with your expectations?

Instructions:
Analyze > Multivariate Methods > Multivariate. Place all 9 perfume variables in the Y, Columns role. Click OK. Focus on just the correlation table for this report.

Scatterplot visualization: Graph > Graph Builder. Place 'Amour Wind' on the Y axis. Choose Redo > Column Switcher under the red triangle. Select all the perfumes including 'Amour Wind' as the Replacement Columns. Now place 'Ivory Reign' on the X axis. Choose Redo > Column Switcher under the red triangle. Choose' Ivory Reign' as the Initial Column to Switch, and all the available perfumes as the Replacement Columns. Now any pairs of perfumes can be selected for the scatter plot.
Note: Because data values that are the same are overlayed on top of each other, choose the Hex Grid jitter in the Points controls to the left of the graph to see the quantity of data for each pair of scores. You can also choose to add a regression line  *to aid in seeing the relationship.*

9. Create a visualization that shows the consumers' scores for each perfume separated by their Age and level of Social Activity. Examine the data for each perfume. Is the consumers' willingness to buy the perfume similar across the age ranges, or does there appear to be an association with age? Briefly describe any association observed. Does there appear to be any significant association in the consumers' scores with level of social activity?

Instructions:
Graph > Graph Builder. Place 'Amour Wind' on the Y axis. Place 'Age' on the X axis and 'Social Activity' as the Group X.
Choose to display both the individual data points and a line graph. 

Choose Redo > Column Switcher under the red triangle. Choose 'Amour Wind' as the Initial Column to Switch. Select all the perfumes including 'Amour Wind' as the Replacement Columns. Now any perfume can be chosen to display.

Create a second version of this graph using 'Social Activity' on the X axis and 'Age' as the Group X.

10. To formally test if there are significant differences, perform a one factor ANOVA for each perfume by amount of Social Activity. Is there significant statistical evidence indicating that the consumers' scores are not the same for these three groups? Is this result consistent with what you observed in the graph created in Exercise 9?

Instructions:
Analyze > Fit Y by X. Place all the perfumes in the Y, Response role. Place 'Social Activity' in the X, Factor role. Click OK. For each perfume, display the ANOVA table by selecting Means/ANOVA under the red triangle next to each graph.
Note: holding the Ctrl key (in Windows) or the Command key (Mac) while selecting an option under the first red triangle, will apply that to all perfumes.

11. Perform a one factor ANOVA for each perfume by Age. Is there significant statistical evidence indicating that the consumers' scores are not the same on average for each age group? Conduct pairwise multiple comparisons across the 8 age groups (28 in total). Is this result consistent with what you observed in the graph created in Exercise 9? Choose two of the perfumes and summarize the results of the multiple comparisons.

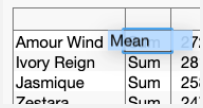
Instructions:
Analyze > Fit Y by X. Place all the perfumes in the Y, Response role. Put 'Age' in the X, Factor role. Click OK. For each perfume, display the ANOVA table by selecting Means/ANOVA under the red triangle next to each graph.
To perform multiple comparisons, choose Compare Means > Each Pair, Student's t under the red triangle.

12. Perform a two factor ANOVA for each perfume by Age and Social Activity. Are the results of the statistical tests corresponding to this more complete model change your conclusions reached in Exercises 10 and 11?

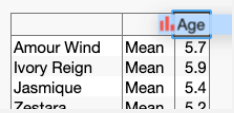
Instructions:
Analyze > Fit Model. Place all the perfumes in the Y role. Select 'Age' and 'Social Activity' in the columns list. Then choose Factorial to Degree under the Macros menu next to the model effects window. Note: This choice will fit a two factor ANOVA model with both main effects and the two-way interaction. Select the Fit Separately check box and choose Minimal Report in the Emphasis menu. Click OK. Examine the p-values in the Effect Summary table for each perfume.

13. Create a new data table that is the average consumer's score for each perform for each age group.

Instructions:
Analyze > Tabulate. Drag all 9 perfume variables into the Drop zone for rows. Select and drag the Mean from the summary statistic list on top of the word Sum in one of the rows in the table to use that as the summary statistic to display.



Select and drag the variable 'Age' and place in the top drop zone above the mean values.



Change the format to display the data to the 10th decimal place by selecting the Change Format button in the bottom left. Select the "Use the same decimal format" check box. Choose Fixed Dec and 1 for Number of decimals.

Choose Make into Data Table from the top red triangle. Save the file using a descriptive name such as 'Avg Consumer Scores by Age'.

14. Create a new data table that merges the average expert evaluators' aroma scores (from Exercise 3) with the average consumers' scores (from Exercise 13).

Instructions:
Open both the data tables ('Avg Expert Evaluators Aroma Scores' and 'Avg Consumers Scores by Age'). Confirm the name of the first column in each table is 'Perfume'. Change if not.
Tables > Join. Select the table to be joined with the current active data table. On the left on the dialog box, select 'Perfume' in the variables list for both data table and click the Match button in the Match Columns section. Click OK.
You'll now have a new data table with both the average experts' and average consumers' scores by 'Age' for each perfume. Save the file using a descriptive name such as 'Avg Expert and Consumer Scores'.
Delete one of the columns with the perfume names, and rename the one saved to "Perfume". It will also be helpful to delete all the saved scripts by right-clicking on them in the scripts window and selecting delete.

Exercises 15-16 will use this new combined data set containing both the experts' and consumers' scores.

15. Create a scatterplot visualization that will allow you to display the average consumer scores for each age group by the average expert scores for each aroma attribute. Choose just two of the consumer age groups and provide an interpretation of their willing to buy a perfume based upon the presence of certain aroma attributes.

Instructions:
Graph > Graph Builder. Place the variable '< 20' on the Y axis. Choose Redo > Column Switcher under the red triangle. Select all the age groups including '< 20' as the Replacement Columns.
Place the variable 'Mean(Sweet)' on the X axis. Choose Redo > Column Switcher under the red triangle. Choose 'Mean(Sweet)' as the Initial Column to Switch, and all the aroma variables as the Replacement Columns. Now any pairs of age group and aroma attribute can be selected for the scatterplot.
Choose to display a linear regression line for the graph  *along with the* R^2 *, Equation, and F Test from the Line of Fit controls on the left.*
The intent of this visualization is to examine any potential relationships between the average consumers' scores for a given age group with the aroma attributes.
Note: There are 8 x 11 = 88 separate relationships that can be examined.

With an age selected, go through the eleven aroma attributes and look to see where there is a positive, negative, or no correlation.

There is a platform called 'Response Screener' under the Analyze > Screening menu. Using all age group variables in the Y, Response role and all the aroma attributes in the X role, 88 separate linear regressions will be fit and a set of visualizations and numeric output created summarizing the results. See Predictive and Specialized Modeling > Response Screening under www.jmp.com/help to learn about this platform.

16. Based upon the analyses you conducted in Exercise 15, why is not necessarily correct to say that if any of the perfumes had more or less of a particular aroma attribute, than the consumers' scores would increase or decrease exactly as the linear regression relationships fit?

17. Provide any ideas you have for additional data, experiments that could be conducted, etc. that would help this company better understand consumers' preference for their perfumes.