

## SENSORY PERFUME STUDY

### RELEVANT JMP PLATFORMS AND STATISTICAL TECHNIQUES

Graph Builder :	Line Graph, Dotplots, Scatterplots, Linear Regression
Distribution :	Histograms, Summary Statistics,
Tabulate :	Creating new data tables of summary statistics
Multivariate :	Correlations, Hierarchical clustering
Fit Y by X :	One-Factor Analysis of Variance, Multiple Comparisons
Fit Model :	Two-Factor Analysis of Variance

### 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 SETS

### Sensory\_Perfume\_Study\_Expert\_Evaluation\_Scores.jmp

Perfume	Names of the nine perfumes
Evaluator	Names of the six expert evaluators
Session	Identifying of the session for each perfume
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)


### Sensory\_Perfume\_Study\_Consumer\_Scores.jmp

ID	ID of the consumer
Age	Age group (< 20, 21-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50+)
Social Activity	Amount of social activity they engage in when they would typically use perfume (Occasional, Moderate, Frequent)
Age_SA	Variable combining the 24 different consumer groups based on Age and amount of Social Activity
Amour Wind	Score quantifying willingness to buy Amour Wind perfume (1-10)
Ivory Reign	Score quantifying willingness to buy Ivory Reign perfume (1-10)
Jasmique	Score quantifying willingness to buy Jasmique perfume (1-10)
Lavora	Score quantifying willingness to buy Lavora perfume (1-10)
Muskara	Score quantifying willingness to buy Muskara perfume (1-10)
Nocturne	Score quantifying willingness to buy Nocturne perfume (1-10)
Oceanéa	Score quantifying willingness to buy Oceanéa perfume (1-10)
Saltflower Veil	Score quantifying willingness to buy Saltflower Veil perfume (1-10)
Zestara	Score quantifying willingness to buy Zestara perfume (1-10)


## EXERCISES

Exercises 1-6 will utilize the data table: *Sensory\_Perfume\_Study\_Expert\_Evaluation\_Scores.jmp*

1. Create a set of line graphs displaying the aroma scores for each perfume by all expert evaluators. Do the evaluators have reasonably similar scores across the attributes, or is there a significant amount of disagreement? Are there any 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:** Launch 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. 

An alternative graph can be made 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 set of graphs 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, 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  $6 \times 2 = 12$  values for each attribute for each perfume.

**Instructions:** Use 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.

	Sweet	Fruity	Flora
Mean	Sum	Sum	Sum
440	373	60	


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

	Sweet	Fr
Mean	Mean	Me
Perfume	0740741	3.4537037

Change the format to display the data to the 10<sup>th</sup> 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 'Expert Evaluation Avg Scores'. Exercises 4-6 will use this new data 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 similar to each other. Note: We will perform a more formal analysis in exercise 6 to group perfumes that are similar.

**Instructions:** Launch 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 will help comparing the perfumes a bit easier, is to use 'Perfume' in the overlay role. Then, 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 by 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:** Launch 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 answer in Exercise 4? If a consumer liked one of the perfumes in a given cluster, what would be your expectation of their liking of the other 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., Jasmique)? Why or why not?

**Instructions:** Launch 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 dendrogram 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 7-12 will utilize the data table: Sensory\_Perfume\_Study\_Consumer\_Scores.jmp


7. Summarize the scores from all 477 consumers for each the 9 perfumes. Do any of 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:** Launch 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 scatterplot visualization that will allow you to visualize the data 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:** Launch 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. Launch 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 scatterplot.

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. Is the consumers' willingness to buy a particular 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:** Use 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 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:** Use Analyze > Fit Y by X. Place all the perfumes in the Y, Response role. Put '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 Means/ANOVA in the first perfume, will run that analysis for 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:** Use 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. Note: holding the Ctrl key (in Windows) or the Command key (Mac) while selecting Means/ANOVA in the first perfume, will run that analysis for all perfumes.  
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:** Use 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 score for each age group for each perfume. Merge this table with the 'Expert Evaluation Avg Scores' table created in Exercise 3.

**Instructions:** Use 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 column drop zone. Change the format to display the data to the 10<sup>th</sup> 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 'Consumer Avg Scores by Age'. Change the title of the column with the perfume name to 'Perfume', and delete the column with the label 'Statistics'.

With both data tables ('Consumer Avg Scores by Age' and 'Expert Evaluation Avg Scores' open, choose 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 Math Columns section. Click OK. You'll now have a new data table that with the avg expert evaluation scores and avg consumer scores combined. Save the file using a descriptive name such as 'Avg Expert and Consumer Scores'. You can choose to delete one of the columns with the perfume names and rename the one you 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.

14. 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 liking of perfumes based upon the presence of the eleven aromas in those perfumes.

**Instructions:** Launch 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. Now 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  and also to display the R<sup>2</sup>, 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 consumers' scores with the aroma attributes. With an age selected, step through the eleven aroma attributes. Do so for all the age groups and aroma attributes.

Note: There are  $8 \times 11 = 88$  separate relationships that can be examined.

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](http://www.jmp.com/help) to learn about this platform.



15. Based upon the analyses you conducted in Exercise 14, 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?
16. 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.