Ingredient Substitutions — Testing a 'Match'

SITARAM DIXIT
Corporate Manager — Fragrances, S.H. Kelkar & Company P. Ltd.
Email: sitaram_dixit@shkelkar.com

For a long time now, no new fragrances are actually created. Fragrance creations are only modifications of earlier creations made either to reduce costs, copy popular competition products, remove or reduce aroma ingredients either due to legislations and guidelines, substitute ingredients due to change in suppliers, or defend the product against ingredient short supply. It will not be incorrect to say that almost half of the product development efforts in the F&F industry today are geared to carry out ingredient substitution and reformulation.

Ingredient substitution and matching exercises may not be glamorous from a perfumer's viewpoint, but are essential to maintain franchisee loyalty and for protecting profit margin. The most important requirement in these cases is not to create or design innovative fragrances, but to maintain *status quo,* i.e. preserve the integrity of the original fragrance profile, albeit at a much lower cost price.

Let us also not forget the fact that reformulation techniques, like any new creative effort, also involves stupendous and voluminous activities and require reliable product evaluation methods so that the resultant products meets the desired quality parameters. In this paper we discuss means and methods that can be used to judge whether the fragrance prepared is close to the original and is in line with customer and consumer response desired, in spite of the change in ingredients that are affected in the product. Analytical tests, descriptive tests analysis and discriminatory difference tests are some tests commonly used to determine if the reformulated product and the control reference standard are truly substitutable or merely acceptable.

Descriptive analysis

In this sensory method, a small group of experts identify and quantify individual product attributes.

Descriptive analysis is very useful in profiling and describing how a new formulation is different from the original reference standard. However, it is not suitable to confirm whether the product formulated is an exact match. This confirmation is possible only by consumer acceptance tests in the final product application.

In any case, the person making an ingredient substitution should not expect to achieve or maintain the exact original sensory profile, as a slight drift is possible and very likely. However, descriptive analysis can be used routinely to help document the original profile and identify any drifting that can take place over a long time period.

Difference test

This test is used to find out whether there is a difference between two products, with respect to any specific attribute or characteristic. Difference test is extremely sensitive and can often indicate minute differences between two similar products.

Various versions of difference test are used in the F&F industry. The most common ones are the Triangle test, Duo-trio test, Dual standard test, Paired difference test, A not A test, and Multiple standard test.

Triangle test

Triangle test is the most popular one and very widely used.

The following example indicates the usefulness of this test.

On the advice of a soap manufacturer client, the fragrance supplier wants to replace Indian sandalwood oil with more economical Australian sandalwood oil due to the rising cost of the Indian variety, in their soap fragrance. The soap manufacturer wants to know whether this ingredient replacement is easily detectable by consumers in the final soap made.

Objective

To determine if soaps made by the use of fragrance containing Australian sandalwood oil, instead of Indian sandalwood oil, can be easily detected by its odour.

Method to use

- The company intends to replace Indian sandalwood oil, one of the main ingredients in the fragrance oil. A small, but sensitive, test is required, before the company undertakes a larger market research study.
- Time is not a constraint; however the test should be reliable.
- Soaps made with both old and new fragrance oils are to be tested.
- A triangle test is most suitable for the purpose. The results can be validated by χ^2 (Chi square) test.

Protocol and Standard operating procedure

- Thirty volunteers having an odour discriminating nose and considered as good odour evaluators are selected as panelists.
- Two soap cakes made with fragrance oil containing Indian sandalwood oil (reference standard / control) are labelled say "K2" and "M8" and one soap cake made out of fragrance containing Australian sandalwood oil (experimental) is marked say "L7".
- The soap samples are all made having

identical formulation (except for the fragrance oil type), shape, size and colour.

- The panellists are asked to smell the soap cake and select the odd one out of the three samples given.

Validity

The function χ^2 is given by

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

where O = observed value and E = expected value.

$$\chi^2 = \frac{(17-10)^2}{10} + \frac{(13-20)^2}{20}$$

$$\chi^2 = 7.35$$

Therefore, the calculated value of χ^2 is 7.35. This value is compared with the table value of χ^2 for given degrees of freedom at a certain specified level of significance (generally, 5% level is selected and used).

If the calculated value of χ^2 is more than the table values of χ^2, the difference between theory and observation is considered to be significant, i.e., it could not have arisen due to fluctuations of simple sampling.

If, on the other hand, the calculated value of χ^2 is less than the table value, the difference between theory and observation is not considered as significant, i.e., it is regarded as due to fluctuations of simple sampling and hence ignored. The results though not statistically significant in the experimental observations may become statistically significant when carried out with larger number of sample observations.

χ^2 are always positive with its upper limit as infinity. As χ^2 are derived from observations, it is a statistic and not a parameter. The χ^2 tests are also termed 'non-parametric' tests, where the test involves no assumption about the form of the original distribution from which the observation comes. The test conducted has one degree of freedom, as there is only one constraint governing the data, namely the total number of selection must be 30.

Conclusion

From the χ^2 distribution table, the value of $\chi^2_{0.05}$ with one degree of freedom, i.e., v = 1, is 3.814. The calculated value of χ^2 is 7.35. Therefore the calculated value $\chi^2_{0.05}$ is higher than the table value. Therefore the difference between theory and experiment is significant and it could not have arisen due to fluctuations of simple sampling. The fragrance odour in the soap made by replacing Indian sandalwood oil with that of Australian sandalwood oil is detectable.

Table 1
Test results obtained

	Correct selection	Incorrect selection	Total
Experimental results	17	13	30
Expected results if no difference found (Hypothesis)	10	20	30

Table 2
Expected frequencies and contingency table

	Correct selection	Incorrect selection	Total
Experimental results	Ox	Oy	N
Expected results, if no difference found (hypothesis)	N / 2	N / 2	N

Duo – trio test

This test is similar to the triangle test. However, in this test the control is identified and presented first. The volunteers are then asked to choose from the remaining sample the one that is different to the control given.

The function χ^2 is given by $\chi^2 = \Sigma (O - E)^2 / E = (Ox - N/2)^2 / N/2 + (Oy - N/2)^2 / N/2$

where O = observed value and E = expected value.

Dual-standard test

In this case, four samples are used. One control and one experimental or different sample are identified to the subjects. Then the subjects are given two unknown samples and asked to pair them. The calculations are similar to the Duo-trio test. Now supposing the control and the experimental samples are not identified in the beginning, but presented together for pairing with two more samples that are similar, the expected frequencies are altered to the standard triangle test. For example, If P_1 and P_2 are the first pair of samples and Q_1 and Q_2 the second pair and if there are no differences between them then the following pairs are possible.

$P_1P_2 : Q_1 Q_2 ; P_1Q_1 : P_2Q_2 ; P_1Q_2 : P_2Q_1$ of which $P_1P_2 : Q_1 Q_2$ is the only correct one.

The function χ^2 is given by $\chi^2 = \Sigma (O - E)^2 / E = (Ox - N/3)^2 / N/3 + (Oy - 2N/3)^2 / 2N/3$

where O = observed value and E = expected value.

Paired difference test

In this case, the control standard is not a part of the test. It is identical to the Duo-trio test.

"A" not "A" test

In this test, the subject is given a control and a test sample and asked to study them, in detail so that they can easily identify the samples. The subjects are then given a series of samples containing both test and control samples. The subjects are then asked to identify and segregate the samples given. The calculation is identical to the Dual standard test method.

Multiple standard test

Sometimes a single product cannot represent the control sample. In such a situation, multiple standard tests are taken recourse to. In this case, a number of samples are used to cover the general product type and presented to the respondent along with the test and experimental sample. The respondent is then asked to pick out the sample that is different from the entire lot of product samples given.

Where E= number of control sample + 1 test sample.

The function χ^2 is given by $\chi^2 = \Sigma\,(Ox - N/E)^2/(N/E) + (Oy - (E-1)N/E)^2/(E-1)(N/E)$

The "R – Test"

This method based on signal detection theory is infrequently used in the F&F industry. The advantage of this method is that it can be used to do multiple comparisons. The R – Index can also measure the degree of difference between two products.

Firstly, the respondent is given a reference control standard. The reformulated product are then individually compared to the reference control and the respondent is questioned "Whether the samples are same or different to the reference control standard". The reference standard is also given to the respondent in blind or coded form as a test sample, in order to measure whether the respondent is able to distinguish the incoming sensory signal (test sample) from background signal (reference standard).

The R – Index ranges from 50 to 100. When R = 100 it means that there is a 100% noticeable difference between test and reference control. When R = 50 it means that a non-distinguishable difference exists between test and control. The higher the R – index the greater the difference between reference standard and sample.

The biggest advantage of "R – test" is that it can be used to rapidly compare a reference sample with several other test samples, all at the same time, using a smaller number of respondents.

Table 3
The contingency table

	Correct selection	Incorrect selection	Total
Experimental results	Ox	Oy	N
Expected results, if no difference found (hypothesis)	N / 3	2N / 3	N

Table 4
The contingency table

	Correct selection	Incorrect selection	Total
Experimental results	Ox	Oy	N
Expected results, if no difference found (hypothesis)	N/E	(E-1)N/3	N

However the disadvantage of R – Index is that, it is not universally accepted as statistical tool for testing, as there are no tables to interpret the statistical significance of the R – Index result, unlike other available standard statistical methods. This is highlighted by the following example:

Popular fragrance spray oil contains nitromusks and other ingredients prohibited by IFRA. In line with IFRA guidelines the fragrance supplier wants to replace nitromusks with polycyclic musk; and other prohibited ingredients with approved aroma chemicals. The spray marketer wants to know whether this ingredient replacement is easily detectable by consumers in the final product.

Objective

To determine whether the fragrance spray made by the use of IFRA-compliant fragrance oil is markedly different in odour profile to the original fragrance spray made with non-IFRA compliant fragrance oil.

Method to use

- Four different perfumers have each created three versions of the original fragrance oil using aroma ingredients permitted by IFRA.
- We have to do multiple comparisons with the basic purpose of measuring the degree of difference between various products, in comparison to standard.
- There are several test samples that have to be tested at the same time with a smaller number of respondents; we use the R – Index method.

Protocol and standard operating procedure

- Ten respondents are presented with reference sample.
- The respondents are then supplied with coded samples of both reference standard control and experimental test sample.
- The panellists are asked to smell the

fragrance spray samples and asked the following questions.

1. Whether they are sure the coded sample is the same as the control.
2. Whether they think, but are not sure, that the coded sample is the same as the control.
3. Whether they think, but are not sure, that the coded sample is not the same as the control.
4. Whether they are sure that the coded samples are different.

Conclusions

The R – Index 86 & 89 are closer to 100. Thus it can be concluded that the samples are different. However, this method does not answer as to how different it is from the standard, although we can say that it is more different than a test sample who's R – Index is say 72. Thus, this method has to be calibrated using products with known differences, so that the R – Index scale has a practical meaning. Similar comparisons can be done with other fragrance test samples and conclusions drawn using R – index. The tests have to be repeated for confirming precision and accuracy of the sensory measurements.

Some more important test considerations

For all descriptive and discriminating difference tests, we need a reference control or standard sample. Selecting a suitable reference standard is by itself a very important parameter for correctly carrying out the analysis. Some important points to be observed while selecting the reference standards include the following:

1. Normally test samples are laboratory prepared. In line with the test samples, the reference sample should also be ideally laboratory prepared. Choosing a sample made in a production plant or using a marketed product as a reference standard, as is normally done, is incorrect and will lead to inconsistent and erroneous test results.
2. It is important that the product on scale up from a laboratory scale to actual plant production is similar. In case this not practically possible, then both standard reference control and experimental test sample should be prepared in the production plant for carrying out descriptive and discriminating difference tests.
3. All tests carried out should be confirmed for consistency of results. This is very much important especially with production or bulk samples, as variation occurring in use of raw materials can have a great impact on test results.
4. Maturity or ageing of the samples before starting the test is also to be taken note of. It is important that the age of the experimental test sample and that of the reference control are same.

Other factors that also need to be considered include as to how close the reformulation is required to be closer to the original to qualify it to be called a "Match". Are the respondents selected experts, semi-experts, regular consumers? This is important as an expert panel can report difference between reference control and experimental test sample as statistically different, that regular consumers might not even remotely notice. It is important that the levels of deviation acceptable should be set in advance prior to starting the tests, so that the results obtained are credible without any bias.

Performance & stability testing

A newly formulated product may clear the sensory evaluation tests, but this by itself is not sufficient to conclude that the ingredient change is successful.

Performance testing in the final application product is necessary. In use, tolerance and shelf-life testing is essential and should be confirmed. Instead of only a simple difference testing for shelf-life, a descriptive analysis with a quality rating approach is preferable, because in a simple test where comparison of the newer formulation with the older one is carried out, it is quite possible that the newer formulation that is actually superior in profile and performance might fail the discriminatory test simply because it is different from the reference original.

CONCLUSION

All the various procedures discussed are straightforward simple methods, regularly used in product testing.

However, it is important that evaluation methods have to be customised for each ingredient substituted in a product so that a sound logical test plan considering the potential problems is designed to know when you have made a "Match".

Table 5
Test results obtained

Total Number of respondents = 10	Sure Signal	Doubtful Signal ?	Doubtful Noise ?	Sure Noise	R – Index = (S / N) X 100
Sample (Sensory Signal)	6	2	2		Level I (6 / 7) X 100 = 86
Standard (Background Noise)		1	2	7	(8 / 9) X 100 = 89 Level II

Additional Reading

1. Consumer Product Evaluation, Principles and Practice, Sitaram Dixit, *STDR Annual 2003*, also *Indian Perfumer* (EOAI Journal), Vol. 49. No. 4, October – December 2005.
2. Is stability testing of consumer products really necessary? Sitaram Dixit, *Chemical Weekly*, August 7, 2001.