# R&D REPORT
## NO. 119

**Proficiency testing for
sensory profile tests:
statistical guidelines – part 1**

**2000**

Campden BRI

R&D Report No. 119

# Proficiency testing for sensory profile tests: statistical guidelines – part 1

JA McEwan

2000

# EXECUTIVE SUMMARY

Proficiency testing in sensory analysis is an important step which aims to demonstrate that data obtained from trained sensory assessors are as reliable as one would expect from any other objective measurement tool. Sensory analysis is unique in that it uses human assessors to measure the perception of a wide range of stimuli, as detected by the senses of sight, sound, smell, taste and touch. Such measurements are physical translations of perception, and as such differ from other physical or chemical measures.

The uniqueness of sensory analysis poses some specific problems for measuring the proficiency of the sensory panel. Cultural and psychological/physiological differences may give rise to differences in perception, and the panel's product experiences may lead to differences in the ability to discriminate between samples. Such factors make the job of the statistician more demanding; defining the expected level of performance in terms of sample discrimination, for example, becomes difficult. Another issue is the definition of a 'true value' or 'expected result', which is not so clearly defined for sensory analysis.

There are a number of methods in the literature that could be used to evaluate the performance of sensory panels for profile tests. Methods based on analysis of variance are widespread, and the use of generalised Procrustes analysis is well documented. These and other methods are investigated for their potential use in proficiency testing, and selected methods are explored further using data collected as part of a ring trial during 1999.

This first stage of research proposes a procedure for the establishment of performance criteria for future ring trials, and how panels can be assessed according to these measures. Moreover, the important issue of 'true value' in proficiency testing seems to have been partly resolved through the calculation of an 'expected result'. However, further work is required to define the procedure for determining the 'expected result'.

# ACKNOWLEDGEMENTS

# CONTENTS

# 1. INTRODUCTION

## 1.1 Background to Proficiency Testing

Proficiency testing in sensory analysis is an important step to demonstrate that data obtained from human instruments are as reliable as one would expect from any other objective measurement tool. Sensory analysis is unique in that it uses human assessors to measure the perception of a wide range of stimuli, as detected through the senses of sight, sound, smell, taste and touch. Such measurements are physical translations of perception, and as such differ from other physical or chemical measures.

The uniqueness of sensory analysis poses some specific problems for measuring the proficiency of the instrument (panel) providing the data. Cultural and individual differences may give rise to differences in perception, and product experience of the panel may lead to differences in the ability to discriminate between samples. Such factors make the job of the statistician more difficult. For example, defining the expected level of performance in terms of which samples are differentiated.

Another issue for the statistical evaluation of the data is the definition of a 'true' value, which is not clearly defined for sensory analysis. In the case of ranking, the most logical definition is the rank order of the samples according to the way in which they were spiked, though more complex spiking may make the task of identifying the true ranking more difficult (McEwan, 2000). However, this issue is even more problematic for descriptive profile data where panels use different attributes, and may not have the same definition of an attribute that has the same name (in English, or any other language).

This document outlines approaches to the analysis of sensory profile data, with the specific objective of monitoring the performance of the panel as part of a sensory proficiency testing scheme.

## 1.2 Methods Covered

There are many methods used to analyse descriptive profiling data, both in relation to receiving information about the samples, but also in relation to evaluating the performance of the panel. Methods covered in this document include analysis of variance (ANOVA), principal component analysis (PCA), generalised Procrustes analysis (GPA), individual differences scaling (INDSCAL) and the RV coefficient.

In considering the role of these methods, one issue that is uppermost is the fact that the sensory panels use different numbers and types of sensory attributes. It is essential to resolve this issue to enable meaningful comparisons of performance to be made.

Each of the methods will be presented with their role in evaluating the performance of a panel. It is the practical application of these tools, that will allow their feasibility to be established and their future role for use in proficiency treating for sensory profile analysis.

## 1.3 Panel Performance or Assessor Performance

One important aspect to clarify at the outset, is the purpose of proficiency testing with respect to performance of panels or performance of assessors.

It is very clear, that whether in research or commercial projects, it is the panel result that is used to make decisions about the samples being evaluated. Therefore, proficiency testing is about measuring the performance of a panel, not individuals within the panel.

If individual assessors perform poorly, then their data will bring down the overall performance of the panel, and therefore the panel will not have performed well. However, the concordance between members of the panel is of interest. It is one measure of a panel's performance that determines the closeness of information provided by each panel member.

In summary, this document is concerned with the performance of panels, and not individual assessors within the panel.

## 1.4 Samples for Proficiency Testing

Another consideration for proficiency testing is the choice of samples to ensure that reliable measures of performance can be defined. There are three possibilities: real foods from the market; 'spiked' samples specially prepared for the test; or model samples.

Model solutions are the easiest to deal with, but are arguably not realistic of a typical profile assessment undertaken by a trained sensory panel. Market samples can be purchased in one country for distribution and offer a ready-made set of products, but information about the sample composition is not normally known in any detail. Finally the use of 'spiking' allows a base product to be used, with control over how the product is modified, but may produce samples which are too simple.

## 1.5 Document Format

Chapter 2 covers the important issue of experimental design, whilst Chapter 3 reports on the standard procedure for analysing descriptive profile data to determine if the samples are significantly different, which samples are different, and on what key attributes.

Chapter 4 outlines the details of the 1999 ring trial, to enable the reader to make best use of the illustrative examples in individual methods chapters. Chapter 5 treats the important subject of 'true value' or expected result, which is central to the success of measuring panel performance for proficiency testing.

Chapter 6 looks at the role of ANOVA, both in relation to the pre-selected common attributes used by each panel, and also with respect to the sensory dimensions derived from different multivariate analysis procedures.

Chapter 7 explores the use of PCA to establish how many sensory dimensions are used to discriminate between samples, and whether these are the same for each panel.

Chapter 8 investigates the extended role of GPA over PCA, but concentrates on the concordance between assessors within a panel and between panels. The role of GPA is similar to PCA, but allows the data to be examined in one analysis. Moreover, this method allows the use of the common attributes to be explored.

Chapter 9 considers the used of INDSCAL for comparing panels, where it is demonstrated that the problem of different attributes across panels may be overcome. Moreover, this method allows the individual dimensionality of the sensory space be compared across panels.

Chapter 10 explains the use of the RV Coefficient for comparing the similarity between multivariate sample maps. This method is analogous to a multivariate correlation coefficient.

Chapter 11 summarises the results for the ring trials, whilst Chapter 12 proposes a scheme to measure performance of panels for future ring trials

# 2. EXPERIMENTAL DESIGN

## 2.1 Introduction

Statistically designed and analysed comparative experiments are of greatest value in those experimental circumstances where treatment (or Sample) effects are likely to be small compared to the underlying variation. Statistical analysis is carried out to estimate the effects of treatment and to assign well-founded estimates of variation to treatment effects. This in turn leads to either tests of significance or to confidence intervals.

In order to satisfy the assumptions of 'statistical' analysis there should be elements of randomisation in the design. In order to increase the precision of the experiment in estimating differences between treatments, it is usual to identify known sources of extraneous variation and to seek to 'block' by these factors i.e. incorporate them into the design of the experiment. For sensory profiling experiments these are Assessor, Order of Sample Presentation and the effects of Previous Sample.

The interpretation of treatment effects is greatly facilitated by the adoption of a factorial treatment structure for the Samples. The precision of an experiment can be improved by increasing the replication.

For sensory science, in general, there has been a recognition of a lack of awareness of the advantages of carefully designing sensory experiments (Hunter, 1996; MacFie et al., 1989) and of developments in the analysis of such data (Jones and Wang, 2000). Sensory profiling experiments, in particular, are very similar statistically to the 'cross over' designs used in pre-clinical and clinical medicine (Jones and Kenward, 1989). These designs were originally used for in vivo animal studies in the biological sciences.

Below, the logic of proposed designs for sensory profiling are developed. It is recommended that such designs are also used in Ranking experiments. This is divided into

three parts, first the Assessor design, second the design of the Samples and thirdly Replication. An alternative fuller account is given by Hunter (1996).

## 2.2 The Assessor Design

It is common to find that Assessors are the largest effect in the Analysis of Variance of the data for each attribute. This arises because Assessors use different parts of the scale. Nevertheless, Assessors provide useful information on the differences between Samples. Provided that each Assessor tests each Sample the same number of times, it is possible to estimate treatment effects entirely within Assessors. Assessors are thus a block factor.

For data from Sensory Profiling experiments, 'Order' effects are also known to be important (Muir and Hunter, 1991/2). A Sample tested first in a Session is usually rated differently from the same Sample tested later in a Session. Such is the magnitude of this effect that it is important to either randomise Order of presentation within a Session or alternatively design it into the experiment as a block effect in addition to Assessors.

The experience indicates that Assessor and Order effects can be effectively designed to be block effects using designs based on Latin Squares. Furthermore, if the cyclic Latin Squares of Williams (1949) are used as a base for the design then protection is provided against interference effects from the Previous Sample (MacFie *et al.*, 1989; Hunter, 1996). Although there is very little evidence to show that these interference effects are important (Muir and Hunter, 1991/2), it is prudent to design the experiment so that treatment estimates are protected. Hunter (1996) shows how Williams Latin Squares can be used to generate statistically efficient yet practical designs for nearly all profiling experiments.

## 2.3 Sample Design

For many experiments it is not possible to impose a factorial treatment structure on the Samples. However, when the Samples are from the laboratory it is usually possible to structure them in a factorial manner. Such structuring improves the ability to interpret the results of the experiment. A complete set of all factorial combinations is often used and can be very helpful in identifying 'active' factors. In those circumstances where the number of factorial combinations is too numerous for one sensory experiment then fractional designs should be considered. It is possible to find designs from the literature which allow the main effects of three factors to be investigated with four Samples, seven factors with eight Samples and eleven factors with twelve Samples.

## 2.4 Replication

Replication is primarily used to increase the precision of the Sample estimates although it should not be used as a substitute for an undersized panel of Assessors. A secondary advantage is that Replication allows each Assessor's reliability (i.e. the agreement between different ratings of the same Sample by the same Assessor) to be determined.

The word Replication is used in a number of different circumstances in Sensory Science and does not appear to have the precise usage that is seen in the Biological Sciences. Below we explain our understanding of Replication and following it we explore an alternative idea about Replication.

### First Scenario

In the first Replicate, the Assessors rate each Sample, if necessary spreading the assessments over a number of Sessions. In the second and subsequent Replicates, the

Assessors rate the Samples again using a new randomisation that preserves the "blindness" of the trial. If Assessors are given Samples in the same order in each Replicate then they will eventually become aware of this fact and will anticipate the results thus nullifying the concept of independent ratings. Also, different randomisations for each Replicate allows each Assessor's data to be independently tested for Order and Session effects.

Where the number of Samples require more than one Session per Replicate for assessment, then it is desirable, on statistical grounds, that in each Replicate, the Samples are divided into the Sessions differently for each Assessor. If this is not possible (i.e. when hot Samples are being tested) then the Samples should be divided into Sessions differently in each Replicate.

### Second Scenario

When more than one Session per Replicate is required, the alternative to the above is to randomly allocate each Replicate of each Sample to a session with the restriction that Replicates of the same Sample cannot occur in the same Session. An incomplete block design (Fully Balanced or Partially Balanced Incomplete Block Design) might be used for this purpose. If at all possible, different randomisations of the design should be used for each Assessor.

Overall, the *"Second Scenario"* offers no advantage and suffers from the disadvantage of requiring the design to be completed in order to yield data that can be easily analysed. The *"First Strategy"* can be recommended as it allows the experiment to be completed one Replicate at a time and because it allows learning effects and/or changes in the Samples to be easily monitored.

Finally, it is desirable (although infrequently realised) that for Replicate assessments, new Samples are drawn for each Replicate. This then allows variability between Samples of the same product or formulation to be incorporated into the experiment. Otherwise specific

Samples are being compared without allowing for sampling or manufacturing variation.

## 2.5  Conclusions

Careful design of sensory experiments, using well established techniques freely available in the literature, allows the maximum amount of information to be derived from the work of the Sensory Laboratory.

# 3. ANALYSIS OF DESCRIPTIVE PROFILE DATA

## 3.1 Introduction

Descriptive profile data can be analysed in a number of ways, including descriptive statistics, analysis of variance (ANOVA) and principal component analysis (PCA). This Chapter outlines the purpose of each of these methods for analysing and interpreting descriptive profile data.

## 3.2 Summary Statistics

These are calculated for each sample, on each of the attributes in the descriptive profile. Whilst numbers can be examined, it is also recommended that suitable graphical methods such as histograms and line graphs are used to explore the data.

The *arithmetic mean* is the most commonly calculated summary statistic, and is the average value of the data. This allows the user to get a feel for which samples are different, and for the intensity of each attribute across the samples being evaluated. The mean data are routinely reported as part of a report on profiling results, and should be consulted throughout the interpretation process.

The *standard deviation* is a common measure of spread of the data around the mean, and is the size of the middle range of the data. Ideally each sample should have a small standard deviation if the panel is operating well. Large standard deviations can be due to one or more assessors using the upper part of the scale compared to other assessors who used the lower part of the scale, or to poor agreement between members of the panel. In addition, large standard deviations can be caused by within sample variation (e.g., due to batch

problems or inconsistent sample preparation). However, this last point should not be the case for proficiency testing samples, where homogeneity is tested in advance.

The *standard error* is a measure of precision of the mean, and is a function of the standard deviation and the number of observations per sample.

It is also useful to record the *range* of the data, which is simply the difference between the *maximum* and *minimum* value given to a sample, for a specific attribute. This allows potential problems to be identified.

## 3.3   Analysis of Variance

As its name suggests, ANOVA breaks down the different sources of variation in the data. Its aim is to establish whether there are differences between the sample means for a specific attribute. In general, a two-way analysis of variance is undertaken, specifying assessors and samples as the two factors, together with the interaction between assessors and samples (replication is required). In general, a mixed model would be used where assessors are considered to be a random effect and samples a fixed effect. This is because analysis of the profile aims to make statements with respect to the samples in relation to differences that may be perceived by a larger population (a pool of sensory assessors). However, a fixed effect ANOVA may be considered more suitable for evaluating the performance of a panel undertaking a proficiency test.

By calculating the assessor effect, this allows the variation due to scaling differences to be eliminated from the residual noise in the data, thus allowing the sample effect to be calculated against the smaller residual error term. When the assessments are replicated an interaction should be specified in the ANOVA table, and the sample effect should be calculated against the interaction. This strategy is consistent with the mixed model approach.

Another consideration for the global ANOVA is whether there is any structure to the samples or not. If there is no structure to the samples, then the ANOVA described above is appropriate. However, if the samples have been created according to a design, then this information should be used in the ANOVA, for example the sample may contain two factors: sugar and acid.

## 3.4   Sample Differences – Multiple Comparison Tests

The ANOVA tables allow each attribute to be investigated with respect to discriminating between samples. However, it does not reveal which samples are significantly different from each other. To investigate this, a suitable statistical multiple comparison methods needs to be performed. Choice of method depends partly on the objective of the experiment. Fisher's least significant difference is widely used and is suitable where a small number of comparisons are made, though it does tend to find differences when the ANOVA result just missed being significant at a specified level. The Tukey's HSD multiple comparison test offers a robust method for larger sample sizes. This topic is discussed further in Chapter 4.

## 3.5   Principal Component Analysis

PCA is a multivariate method that allows the profile data to be summarised in a smaller number of dimensions than the total number of attributes in the profile. In general, between two and four new dimensions are sufficient to summarise the data, but this will depend on the complexity of the samples and the ability of the assessors to use the vocabulary to measure differences between samples.

The PCA (covariance option) is normally undertaken on the mean profile data, after

removing outlying observations in the data. This is because it is important to retain the original variance structure of the attributes. If the correlation option of PCA was chosen, then all attributes would be rescaled, and this could result in unimportant (non-discriminatory attributes) being highlighted as important.

As well as undertaking PCA on all the profile attributes, a separate PCA may be undertaken on the attributes of each of the sensory categories examined (e.g. appearance, flavour, texture). For the purposes on this project, means will be calculated on the whole data set.

## 3.6  Generalised Procrustes Analysis

GPA is a multivariate method that considers differences in the way assessors use the scales to measure the sensory attributes, and the way in which these attributes are used.

The principle of GPA is similar to that of PCA, except 3 key steps are undertaken on each assessor's profile matrix (samples by attributes) prior to deriving a consensus output. Translation to a common origin adjusts for assessors using different levels of the measurement scale, whilst isotropic scaling adjusts for assessors using different ranges of the measurement scale. The step of rotation and reflection adjusts for assessors using attributes in a different way, and is an important step in establishing the performance of the panel.

The final output of GPA is the same as that from a PCA, a graphical representation of samples and attributes in several dimensions. One important aspect of GPA, is it allows the user to determine if assessors have used each attribute in the same way, and if not, which assessors are different.

# 4. INFORMATION REGARDING 1999 RING TRIAL

## 4.1 Sample Information

Six samples of tomato soup were produced according to an experimental design, as shown in the Table 4.1.

**Table 4.1**: Tomato soup samples used in the first ring trial.

| Sample | Product | Cornflour | Flavour | Code |
|--------|---------|-----------|---------|------|
| TS-11 | 904 (77) | No | None | Base |
| TS-12 | 618 (38) | No | Low | F1 |
| TS-13 | 250 (70) | No | High | F2 |
| TS-21 | 591 (32) | Yes | None | C1 |
| TS-22 | 219 (95) | Yes | Low | C1F1 |
| TS-23 | 315 (46) | Yes | High | C1F2 |

## 4.2 Assessors and Attributes

Table 4.2 shows the number of assessors and attributes for each panel. Five common attributes were included by each panel: overall strength of odour, total strength of flavour, thickness, sweetness and saltiness.

Table 4.3 shows the scales used by each of the panels. A formula to convert these to range from 0 to 100 is provided in Appendix 1.

**Table 4.2**:     Number of assessors and attributes for each panel.

| Panel | Number of Assessors | Number of Attributes |
|-------|---------------------|----------------------|
| N | 15 | 20 |
| P | 9 | 17 |
| Q | 7 | 13 |
| R | 8 | 15 |
| T | 16 | 24 |
| U | 11 | 13 |
| V | 8 | 30 |
| W | 10 | 17 |
| Y | 8 | 15 |
| Z | 9 | 17 |

**Table 4.3**:     Scales used by the ten panels.

| Panel | Scale | | Panel | Scale |
|-------|-------|--|-------|-------|
| N | 1 - 9 | | U | 1 – 9 |
| P | 0 – 100 | | V | 0 – 15 |
| Q | 0 – 100 | | W | 1 – 9 |
| R | 0 – 100 | | Y | 0 – 10 |
| T | 1 -10 | | Z | 0 – 10 |

# 5. THE EXPECTED RESULT

## 5.1 What is the True Value

The 'true value' or 'true result' issue is problematic for sensory analysis, and this is particularly so for the descriptive profile. In general, no two panels will use the same attributes, and where the same word is used, it could have a different meaning. This is particularly true where panels use different languages. For everyday sensory analysis this is not important, as it is the ability of the panel to provide meaningful information about the samples, and this information should be derived from any well trained sensory panel.

## 5.2 Establishing the Expected

Rather than considering the concept of the true result, it is more correct to think in terms of an expected result (or results) from a sensory profile exercise. This can be thought of as the nature and size of sensory differences that one would expect a good sensory panel to detect. Of course, this may not be known in advance, and therefore the setting down of the expected results needs careful thought.

There are essentially two main scenarios for selected samples to undertake a sensory profiling. The first is where the samples are structured in some was through spiking (e.g. using a factorial design on 3-7 ingredients). In this situation, the structural information should be recoverable. Careful screening is desirable, and sample selection is discussed in the Proficiency Testing Guidelines (Lyon, 2001).

The second scenario is where samples are selected (from the market) to reflect a range of sensory characteristics. Even though some screening will have been required, the expected result may be considered more difficult to establish. However, even with structured

samples, unexpected interactions can take place between ingredients.

Discussions about this issue continue. The derivation of sensory dimensions, through an appropriate multivariate procedure, seems to offer the best starting point for identifying the expected result. This is because each panel uses their own attributes to describe the samples.

Given that these sensory dimensions cannot be known in advance, unless previous data are available, the expected result can be derived as a consensus from all the data provided in the ring trial. However, there must be a sufficient number of panels to have confidence in this consensus. Furthermore, no one panel can be allowed to distort the consensus. Moreover, in a ring trial, the panels could feasibly all be good or bad. Therefore, the consensus approach does not seem to offer a solution.
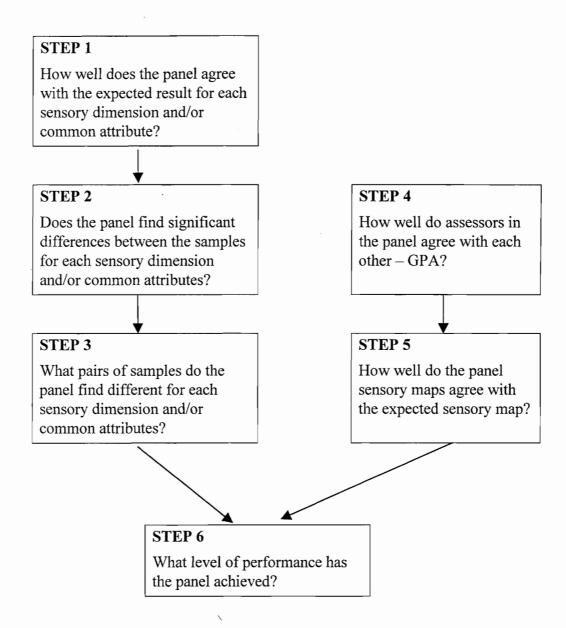
Another approach is to undertake pre-testing with a number of selected panels (3-4), including a panel who has previous demonstrable expertise with the product. In this way, pre-test data can be examined, and expected sensory dimensions defined.

Once this has been achieved, the expected order of samples along each sensory dimensions can be stated as the *expected sample order*.


## 5.3   Establishing Expected Discrimination Between Samples

The definition of sensory dimensions now needs to be seen in light of how well the dimension discriminates between the samples, and further what samples are detected as significantly different. The first step is achieved by analysis of variance, and the second through the use of a suitable multiple comparison procedure.

## 5.4 Stages in Establishing Panel Performance

**STEP 1**

How well does the panel agree with the expected result for each sensory dimension and/or common attribute?

**STEP 2**

Does the panel find significant differences between the samples for each sensory dimension and/or common attributes?

**STEP 4**

How well do assessors in the panel agree with each other – GPA?

**STEP 3**

What pairs of samples do the panel find different for each sensory dimension and/or common attributes?

**STEP 5**

How well do the panel sensory maps agree with the expected sensory map?

**STEP 6**

What level of performance has the panel achieved?

## 5.5 The Role of Common Attributes

There has been considerable debate about the inclusion of common attributes. Common attributes should be defined so that they are understood by all panels, thus allowing univariate measures of panel performance to be used. However, problems may exist with defining these attributes. This Proficiency Testing project considered the use of common attributes, and one of the objectives was to determine the advantages and disadvantages.

# 6. ANALYSIS OF VARIANCE

## 6.1 Principle

Analysis of variance (ANOVA) is a widely used tool for investigating whether each sensory attribute, in turn, significantly discriminates between the samples. Panels that perform well show good discrimination.

The actual format of the ANOVA will depend on whether single sensory attributes are analysed or alternatively sensory dimensions. Moreover, the procedure will change if information about the sample structure is known, and can be used as part of the ANOVA procedure.

## 6.2 Procedures

**ANOVA on Single Sensory Attributes**

Analysis of variance can be performed on each attribute of a panel's data, and the mean square error (MSE) and sample F-ratio and p-value recorded. The MSE measures the amount of noise in the data, and the larger this becomes the less likely the panel is to detect differences between the samples. The F-ratio is the ratio between the mean square value for the sample and the MSE, and this is used to calculate the p-value.

Ideally, a panel will have a low p-value (significant discrimination) and a low MSE. However, it is possible to record a low p-value and a high MSE; this would suggest that whilst discriminating between the samples, the panel was not as repeatable as one would like. If both the MSE and p-value are high, then the panel demonstrated an inability to discriminate.

## ANOVA using Additional Information

The ANOVA used above is a simple two way ANOVA with interaction, where assessors are normally considered as a random effect in the model. However, it is desirable to record other information in the data worksheet, such as Replicate and Order of Presentation. This information can be used in the analysis to obtain additional information about the performance of the panel.

The investigation of Order effects is somewhat more difficult due to the unbalanced nature of the data. It is not usual for each sample to be presented in each order to each assessor. To get round this statistical problem, a more general method known as REML (Residual Maximum Likelihood) is used to fit a mixed model to the data. This is a statistical method for analysing categorised data that are subject to variation at different levels within an experiment.

More information on this method, and mixed model ANOVA can be found in the Technical Appendix. This approach will be investigated in further ring trials.

## ANOVA on Sensory Dimensions

On deriving sample scores on new sensory dimensions from a multivariate procedure, it is possible to undertake analysis of variance. However, the replication should be included in the multivariate analysis. In other words, if 6 samples were evaluated 3 times, 18 observations would be included in the multivariate analysis. A one way analysis of variance, specifying sample effect would be appropriate. However, if a design was used to create the samples (e.g. factorial), then the sample factors and interaction(s) could be used in the ANOVA to partition the sample sum of squares and degrees of freedom.

## Use of Multiple Comparisons

If a panel is performing well, it should be possible to identify expected differences between samples. This is achieved using a multiple comparison procedure, such as the Tukey's HSD. This method is suitable for both simple and more complex comparisons (O'Mahony, 1986; Lea *et al.*, 1997). This calculation can be undertaken by most statistical packages.

As an example, consider the data on Total Flavour for Panel Y. A two-way analysis of variance was undertaken, where the interaction between assessors and sample was specified. The analysis of variance table for total flavour is given below (Table 6.1), and a Tukey multiple comparison value of 0.5 was calculated.

**Table 6.1**: Analysis of variance of total flavour for Panel Y.

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Assessor | 7 | 109.417 | 15.631 | 55.65 | 0.000 |
| Product | 5 | 6.0792 | 1.2158 | 4.33 | 0.004 |
| Assessor*Product | 35 | 9.8308 | 0.2809 | 0.94 | 0.566 |
| Error | 96 | 28.6133 | 0.2981 | | |
| Total | 143 | 153.94 | | | |

This Tukey value can then be used to identify what samples are significantly different from each other, at the 5% level of significance (Table 6.2). Only Products 219 and 904 and Products 315 and 904 were significantly different.

**Table 6.2**: Total flavour product means for Panel Y, illustrating the samples that are significantly different.

| Product | Mean | Differences | |
|---|---|---|---|
| 219 | 5.4 | \| | |
| 315 | 5.4 | \| | |
| 250 | 5.3 | \| | \| |
| 591 | 5.3 | \| | \| |
| 618 | 5.1 | \| | \| |
| 904 | 4.8 | | \| |

Alternatively, analysis of variance can be undertaken using the fact that there is structure in the samples. Table 6.3 shows the ANOVA table, whilst the Tukey multiple comparison results are provided in Table 6.4 for flavour as it had 3 levels associated with it (the comparison value is 0.2).

**Table 6.3**:  Analysis of variance of total flavour for Panel Y, using the ingredient information.

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Assessor | 7 | 109.4167 | 15.6309 | 52.44 | 0.000 |
| Cornflour | 1 | 2.6136 | 2.6136 | 8.77 | 0.0039 |
| Flavour | 2 | 2.0617 | 1.0308 | 3.46 | 0.0355 |
| Assessor*Cornflour | 7 | 5.0231 | 0.7176 | 2.41 | 0.0258 |
| Assessor*Flavour | 14 | 1.4717 | 0.1051 | 0.35 | 0.9842 |
| Cornflour*Flavour | 2 | 1.4039 | 0.7019 | 2.36 | 0.1003 |
| Assessor*Corn*Flav | 14 | 3.3361 | 0.2383 | 0.80 | 0.6676 |
| Error | 96 | 28.6133 | 0.2981 | | |
| Total | 143 | 153.94 | | | |

**Table 6.4**:  Total flavour product means for Panel Y, illustrating the flavour levels that are significantly different.

| Product | Mean | Differences |
|---|---|---|
| High | 5.4 | \| |
| Medium | 5.2 | \| \| |
| None | 5.1 | \| |

# 7. PRINCIPAL COMPONENT ANALYSIS

## 7.1 Principle

Principal component analysis (PCA) is one of the most common multivariate methods used to produce a sample and attribute map to describe the similarities and differences between samples. If the panel is performing well, the PCA map should demonstrate good discrimination between samples on a number of dimensions. In addition, there should be no significant effect of replication on the sample scores.

## 7.2 Procedure

The panel data are averaged over assessors to provide a number of rows corresponding to number of samples x number of replicates (e.g. 6 samples x 3 replicates = 18 observations per attribute). A covariance PCA is conducted on these data and a sample map produced. For these, replicate positions of each sample are joined together to form a triangle (in the case of 3 replicates). Close replicate positions reveal good agreement by the panel with respect to the samples position in the sample map, whilst the contrary illustrates poor performance. A panel can show a good result on one dimension, but not on another.

To illustrate the process, Table 7.1 shows the profile data for Panel Y, averaged across assessors. A covariance PCA is then performed on these data to obtain the sample map shown in Figure 7.1 (see Table 4.1 for product definitions).

One way analysis of variance was then undertaken on the first four principal components to determine which provided information about discrimination between the samples. Table 7.2 shows the p-value for the sample effect on the first 4 PCs, together with the percentage variation explained by these sensory dimensions.

**Table 7.1:** Profile data for Panel Y, average over assessors.

| Prod | Rep | Attributes | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|      |     | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 219 | 1 | 4.7 | 3.1 | 2.0 | 2.3 | 2.8 | 6.1 | 5.3 | 3.8 | 3.5 | 2.7 | 2.5 | 2.5 | 2.6 | 2.6 | 3.0 |
| 219 | 2 | 4.7 | 2.9 | 1.9 | 2.8 | 2.8 | 6.2 | 5.3 | 3.9 | 3.9 | 2.5 | 2.3 | 2.4 | 2.9 | 2.6 | 3.7 |
| 219 | 3 | 4.4 | 2.6 | 1.9 | 2.9 | 2.4 | 7.3 | 5.5 | 4.2 | 4.0 | 2.4 | 2.2 | 2.8 | 2.9 | 2.9 | 3.7 |
| 250 | 1 | 4.2 | 2.9 | 2.2 | 2.7 | 2.2 | 3.7 | 5.5 | 3.7 | 4.0 | 2.7 | 2.4 | 2.8 | 3.3 | 3.0 | 4.0 |
| 250 | 2 | 4.4 | 3.1 | 2.1 | 2.3 | 2.4 | 2.8 | 5.3 | 3.1 | 4.0 | 2.2 | 2.4 | 1.9 | 3.5 | 2.8 | 4.3 |
| 250 | 3 | 4.7 | 3.1 | 2.0 | 2.7 | 2.8 | 3.2 | 5.3 | 3.0 | 4.0 | 2.8 | 2.4 | 2.5 | 3.1 | 2.8 | 3.8 |
| 315 | 1 | 4.6 | 3.2 | 2.4 | 2.5 | 2.5 | 6.7 | 5.7 | 3.7 | 4.2 | 2.7 | 2.4 | 2.5 | 3.7 | 2.7 | 4.1 |
| 315 | 2 | 4.9 | 3.4 | 2.2 | 2.6 | 2.7 | 6.5 | 5.4 | 3.6 | 3.9 | 2.6 | 2.2 | 2.4 | 3.5 | 2.8 | 4.0 |
| 315 | 3 | 4.4 | 2.8 | 2.0 | 2.1 | 2.4 | 6.6 | 5.1 | 3.3 | 3.7 | 2.5 | 2.3 | 2.3 | 3.1 | 2.8 | 3.7 |
| 591 | 1 | 4.4 | 3.0 | 2.3 | 2.5 | 2.5 | 7.3 | 5.3 | 4.0 | 3.9 | 2.9 | 2.3 | 2.4 | 3.3 | 2.8 | 4.1 |
| 591 | 2 | 4.6 | 3.3 | 2.2 | 2.3 | 2.6 | 6.5 | 5.3 | 3.5 | 3.8 | 2.5 | 2.3 | 2.4 | 3.1 | 2.6 | 3.3 |
| 591 | 3 | 4.6 | 3.0 | 2.0 | 2.6 | 2.8 | 7.1 | 5.4 | 4.0 | 3.7 | 2.5 | 2.4 | 2.7 | 2.9 | 2.6 | 3.6 |
| 618 | 1 | 4.4 | 3.1 | 2.0 | 2.4 | 2.6 | 3.7 | 5.1 | 3.3 | 3.6 | 2.2 | 2.1 | 2.4 | 3.1 | 2.8 | 3.8 |
| 618 | 2 | 4.6 | 2.6 | 2.1 | 2.3 | 3.1 | 4.0 | 5.1 | 3.8 | 3.3 | 2.4 | 1.9 | 2.4 | 2.9 | 2.7 | 3.5 |
| 618 | 3 | 4.1 | 2.5 | 2.0 | 2.4 | 2.9 | 3.3 | 5.1 | 3.5 | 3.8 | 2.4 | 2.4 | 2.1 | 2.8 | 2.7 | 3.6 |
| 904 | 1 | 4.5 | 3.1 | 2.2 | 2.3 | 2.0 | 3.8 | 4.6 | 3.4 | 3.3 | 2.3 | 2.5 | 2.4 | 2.5 | 2.3 | 3.5 |
| 904 | 2 | 4.4 | 2.9 | 2.0 | 2.3 | 2.1 | 3.7 | 4.8 | 3.3 | 3.4 | 2.2 | 2.0 | 2.1 | 3.0 | 2.1 | 3.6 |
| 904 | 3 | 4.2 | 2.3 | 1.8 | 2.3 | 2.8 | 4.3 | 5.1 | 3.5 | 3.5 | 2.3 | 2.2 | 2.3 | 2.9 | 2.6 | 3.7 |



Figure 7.1: Panel Y - PCA Biplot - Tomato Soup
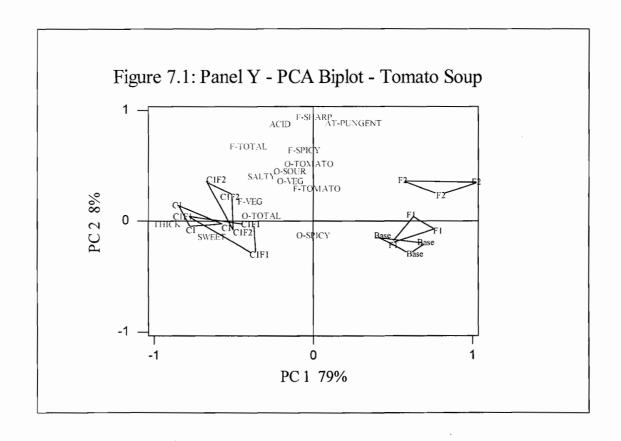
**Table 7.2:** Percentage variation explained by the first four principal components from Panel Y's data, together with the p-value illustrating the sample discrimination demonstrated by these dimensions.

| PC | % | p-value |
|----|------|---------|
| 1 | 78.6 | 0.000 |
| 2 | 8.1 | 0.006 |
| 3 | 4.4 | 0.153 |
| 4 | 2.5 | 0.770 |

Table 7.2 reveals that the first two PCs provide information regarding sample discrimination (see Figure 7.1).

# 8. CALCULATING A CONSENSUS: GPA APPROACH

## 8.1 Principle

A key issue in calculating a meaningful consensus is the differences between panels in the number and names of attributes used, and differences in the portion of the scale used. This situation is, in some ways, analogous to free-choice profiling. GPA provides a means of pulling together data sets, and providing a meaningful consensus in terms of discriminating sensory dimensions. The RV coefficient (Schlich and Guichard, 1989) can be used to measure the correlation between each panel's result and the global consensus.

Similarly, GPA can be used to analyse the data from a panel. It allows differences between assessors in vocabulary and scale use to be accommodated. This is contrary to PCA on the sample means, which assumes that assessors use the vocabulary and scale in the same way.

## 8.2 Procedure

There are a number of approaches to using GPA for proficiency testing, but common to all is the submission of separate data matrices. For example, at a panel level, the sample by attribute matrix for each assessor is submitted to GPA to establish the performance of each assessor in relation to the consensus. Moreover, the use of attributes by each assessor can be investigated to ensure they are being used in a similar way.

The consensus over panels can be established by taking the means for each panel. The panel means are then submitted to GPA to enable the performance of each panel in relation to the consensus to be measured. This exercise can also be undertaken on the mean panel data for the common attributes, thus establishing whether these are used in a similar way.

The correlation (similarity) between the global and panel sample maps can be calculated

using the RV coefficient (see Chapter 10).

To illustrate the GPA, the common attribute data from three panels (Q, T and Y) were submitted to GPA. The data are shown in Table 8.1.
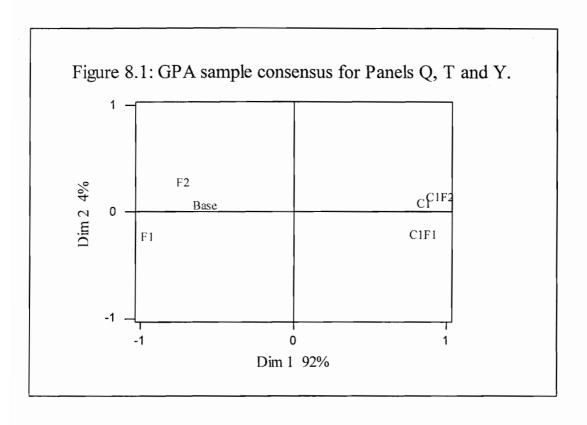
**Table 8.1**: Profile data for Panels Q, T and Y, average over assessors and replicates.
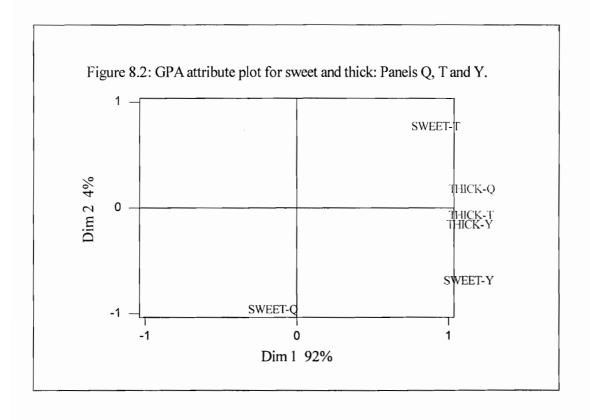
| Panel | Product | O-TOTAL | F-TOTAL | SWEET | SALTY | THICK |
|-------|---------|---------|---------|-------|-------|-------|
| Q | 219 | 60.8 | 64.0 | 53.1 | 51.7 | 45.6 |
| Q | 250 | 57.5 | 61.4 | 50.1 | 53.0 | 37.8 |
| Q | 315 | 56.9 | 61.0 | 49.3 | 52.6 | 47.0 |
| Q | 591 | 55.3 | 60.7 | 53.8 | 54.7 | 44.7 |
| Q | 618 | 61.0 | 62.3 | 55.8 | 49.9 | 33.4 |
| Q | 904 | 59.4 | 62.4 | 52.5 | 52.7 | 37.8 |
| T | 219 | 5.1 | 6.0 | 5.2 | 4.4 | 6.7 |
| T | 250 | 4.9 | 5.7 | 5.2 | 4.4 | 3.5 |
| T | 315 | 5.0 | 5.8 | 5.2 | 4.5 | 6.6 |
| T | 591 | 5.2 | 6.0 | 5.5 | 4.1 | 6.4 |
| T | 618 | 5.0 | 5.4 | 4.6 | 4.1 | 3.4 |
| T | 904 | 4.9 | 5.6 | 5.3 | 4.2 | 3.7 |
| Y | 219 | 4.6 | 5.4 | 4.0 | 2.5 | 6.5 |
| Y | 250 | 4.4 | 5.3 | 3.3 | 2.6 | 3.2 |
| Y | 315 | 4.7 | 5.4 | 3.5 | 2.6 | 6.6 |
| Y | 591 | 4.6 | 5.3 | 3.9 | 2.6 | 7.0 |
| Y | 618 | 4.4 | 5.1 | 3.5 | 2.4 | 3.7 |
| Y | 904 | 4.4 | 4.8 | 3.4 | 2.3 | 3.9 |

*0 = Odour; F = Flavour*

After submitting these data to GPA, a consensus sample map is produced (Figure 8.1). It is now possible to establish the similarity between each panel and the consensus using the RV coefficient. This is described further in Chapter 10.

In addition, it is possible to examine the interpretation of the common attributes. Figure 8.2 shows the position of sweet and thickness for the three panels. It can be observed that there is good agreement on thickness, but Panel Q differs from Panels T and Y with respect to sweet (Dimension 1).

Figure 8.1: GPA sample consensus for Panels Q, T and Y.



Figure 8.2: GPA attribute plot for sweet and thick: Panels Q, T and Y.

# 9.  INDSCAL

## 9.1  Principle

One problem with multi-panel profile studies is that different panels will describe different sensory attributes, and that some panels will use more attributes than others.  This can cause a potential problem in judging the overall performance of the panel.  One way round this is may be the use of the multidimensional scaling method of INDSCAL.

INDSCAL is a well known and accessible multidimensional scaling technique that can be used to fit sophisticated statistical models to sensory data.  In particular it allows differences between assessors or differences between panels to be tackled in an imaginative way.

Within a panel, even when the assessors use the same vocabulary, they may use the descriptors in different ways.  For example, it is not uncommon for assessors to differ in their use of the terms acid, sour and bitter.  Assessors also differ in their use of the part of the scale used (location) and in the amount utilised (scale).  Simple analysis of variance allows for location effects, but not for differences in the amount of scale used or confusion between descriptors.  PCA of means also assumes that assessors use the same proportion of the scale and do not confuse descriptors.  GPA allows assessors to use descriptors in different ways and for assessors to use different proportions of the scale.  INDSCAL allows assessors to weight different sensory dimensions differently by using a separate scaling factor for each sensory dimension.  At the ring trial level, panels will usually have used different sensory vocabularies of different degrees of complexity, and in principle, INDSCAL allows these problems to be tackled.

INDSCAL is a multidimensional scaling technique that works on matrices of differences between samples (dissimilarities).

Multidimensional scaling techniques work on matrices of differences between samples (dissimilarities). Matrices of dissimilarities are akin to the triangular arrays of distances between cities often found in road atlases. Just as the distance between cities can be used to produce a map of the relative location of the cities, the matrix of differences between samples can be analysed by statistical methods to produce a map of the samples in two, three or more dimensions.

Formally, the distance $d_{ij}$ between samples $i$ and $j$ is defined as

$$d_{ij} = \left[ \sum_{a=1}^{r} (x_{ia} - x_{ja})^2 \right]^{1/2}$$

Where $x_{ia}$ specifies the position (coordinate) of point $i$ on dimension $a$. The $x_{ia}$ and $x_{ja}$ are the sample mean values, and $r$ is the number of attributes.

The goodness-of-fit is judged by the "STRESS" statistic. The meaning of the sensory dimensions is found by correlating them against the original data.

When there are several matrices of dissimilarities (say from different assessors) the problem becomes more complicated. A simple solution is to average the matrices of dissimilarity before scaling. A more sophisticated technique is to weight the matrices differently depending on their agreement with the consensus. All such methods presume that there is a consensus. A further level of complexity is to allow assessors to have different sensitivities in each direction (INDSCAL).

The INDSCAL technique, which is a very specialised statistical technique, is available in the program SPSS™ and elsewhere.

## 9.2  Procedure

The data used to illustrate this method come from Panel Y, one of the laboratories who participated in a ring trial profiling tomato soup. Eight assessors used 15 attributes to evaluate 6 products (samples) in triplicate.

**Table 9.1**:     Profile data for Assessor 1 from Panel Y.

| Prod | Rep | Attributes | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 219 | 1 | 4.8 | 2.0 | 3.4 | 2.4 | 1.6 | 1.1 | 5.8 | 3.0 | 3.2 | 2.8 | 3.2 | 1.8 | 3.4 | 0.6 | 0.4 |
| 219 | 2 | 4.4 | 3.4 | 2.8 | 1.4 | 1.6 | 4.8 | 5.4 | 3.0 | 4.2 | 3.6 | 3.0 | 2.4 | 4.1 | 0.6 | 2.6 |
| 219 | 3 | 4.8 | 2.8 | 3.6 | 2.8 | 1.2 | 6.4 | 6.6 | 3.8 | 3.3 | 2.0 | 3.8 | 1.8 | 4.4 | 1.6 | 3.2 |
| 250 | 1 | 5.0 | 2.8 | 3.0 | 3.2 | 0.2 | 1.6 | 5.0 | 3.2 | 3.6 | 2.2 | 3.4 | 1.4 | 3.0 | 0.4 | 0.6 |
| 250 | 2 | 3.4 | 1.6 | 2.4 | 2.2 | 0.6 | 1.4 | 6.8 | 4.4 | 3.4 | 1.0 | 3.2 | 1.0 | 5.0 | 1.8 | 3.4 |
| 250 | 3 | 3.2 | 1.4 | 2.0 | 1.6 | 1.6 | 1.0 | 5.6 | 1.8 | 3.6 | 2.7 | 2.8 | 1.4 | 3.4 | 0.8 | 3.6 |
| 315 | 1 | 4.6 | 3.2 | 3.2 | 1.0 | 0.4 | 5.0 | 6.4 | 2.6 | 3.4 | 2.6 | 2.4 | 1.2 | 4.2 | 0.4 | 1.4 |
| 315 | 2 | 5.4 | 4.0 | 3.4 | 2.6 | 1.6 | 4.9 | 6.4 | 2.0 | 3.4 | 2.0 | 3.2 | 2.2 | 3.8 | 1.4 | 2.4 |
| 315 | 3 | 4.5 | 2.8 | 3.2 | 2.6 | 0.6 | 6.8 | 5.8 | 3.4 | 3.8 | 1.6 | 3.0 | 2.2 | 4.2 | 0.6 | 1.8 |
| 591 | 1 | 4.8 | 1.8 | 3.0 | 2.2 | 0.4 | 7.1 | 6.2 | 3.8 | 4.4 | 3.0 | 4.0 | 2.6 | 4.0 | 1.2 | 2.2 |
| 591 | 2 | 5.5 | 4.0 | 2.6 | 2.4 | 0.6 | 6.6 | 6.8 | 3.8 | 4.6 | 3.2 | 3.2 | 2.8 | 5.4 | 0.6 | 0.6 |
| 591 | 3 | 5.0 | 2.9 | 3.4 | 1.6 | 0.8 | 8.0 | 5.2 | 4.6 | 3.6 | 2.2 | 3.4 | 2.6 | 3.1 | 0.2 | 0.2 |
| 618 | 1 | 5.1 | 2.2 | 2.6 | 2.2 | 0.6 | 2.8 | 5.2 | 2.0 | 2.8 | 1.6 | 3.8 | 1.6 | 1.6 | 1.0 | 1.2 |
| 618 | 2 | 4.6 | 1.0 | 3.8 | 2.2 | 1.6 | 1.4 | 5.2 | 4.0 | 2.4 | 1.2 | 3.0 | 1.6 | 2.4 | 1.4 | 1.2 |
| 618 | 3 | 3.8 | 1.6 | 2.5 | 2.2 | 1.0 | 1.8 | 5.0 | 2.8 | 2.6 | 2.0 | 3.1 | 1.0 | 1.6 | 0.6 | 1.0 |
| 904 | 1 | 5.4 | 3.6 | 3.2 | 2.6 | 1.2 | 1.6 | 5.0 | 3.4 | 3.0 | 3.0 | 2.8 | 2.0 | 2.4 | 0.4 | 0.6 |
| 904 | 2 | 4.2 | 3.2 | 2.8 | 2.2 | 0.4 | 2.2 | 4.8 | 1.8 | 3.6 | 0.8 | 2.5 | 2.0 | 4.0 | 0.6 | 3.2 |
| 904 | 3 | 3.8 | 1.0 | 2.7 | 2.6 | 0.8 | 1.8 | 5.2 | 4.0 | 3.1 | 1.0 | 3.0 | 1.2 | 2.4 | 0.4 | 0.8 |

**Table 9.2**:     Profile data for Assessor 1 from Panel Y averaged over replicates.

| Prod | Attributes | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 219 | 4.7 | 2.7 | 3.3 | 2.2 | 1.5 | 4.1 | 5.9 | 3.3 | 3.6 | 2.8 | 3.3 | 2.0 | 4.0 | 0.9 | 2.1 |
| 250 | 3.9 | 1.9 | 2.5 | 2.3 | 0.8 | 1.3 | 5.8 | 3.1 | 3.5 | 2.0 | 3.1 | 1.3 | 3.8 | 1.0 | 2.5 |
| 315 | 4.8 | 3.3 | 3.3 | 2.1 | 0.9 | 5.6 | 6.2 | 2.7 | 3.5 | 2.1 | 2.9 | 1.9 | 4.1 | 0.8 | 1.9 |
| 591 | 5.1 | 2.9 | 3.0 | 2.1 | 0.6 | 7.2 | 6.1 | 4.1 | 4.2 | 2.8 | 3.5 | 2.7 | 4.2 | 0.7 | 1.0 |
| 618 | 4.5 | 1.6 | 3.0 | 2.2 | 1.1 | 2.0 | 5.1 | 2.9 | 2.6 | 1.6 | 3.3 | 1.4 | 1.9 | 1.0 | 1.1 |
| 904 | 4.5 | 2.6 | 2.9 | 2.5 | 0.8 | 1.9 | 5.0 | 3.1 | 3.2 | 1.6 | 2.8 | 1.7 | 2.9 | 0.5 | 1.5 |

Using the formula given above, the mean data (Table 9.2) were converted into matrices of dissimilarity (Table 9.3). This can be undertaken using programs such as Genstat, S-Plus, SAS and Minitab.

**Table 9.3**:     Dissimilarity matrix for Assessor 1 from Panel Y.

| 219 | 0 | | | | | |
| 250 | 3.41 | 0 | | | | |
| 315 | 2.05 | 4.79 | 0 | | | |
| 591 | 3.70 | 6.67 | 2.83 | 0 | | |
| 618 | 3.84 | 2.92 | 4.91 | 6.55 | 0 | |
| 904 | 3.18 | 2.13 | 4.25 | 6.11 | 1.89 | 0 |
| | 219 | 250 | 315 | 591 | 618 | 904 |

The dissimilarity matrix for each assessor is then read into an INDSCAL program (e.g. SPSS™, SAS™). The user has to specify some input options, and for SPSS™ the measurement level should be *ratio* and the *individual differences Euclidean distance* scaling model should be used. In addition, both the 2 and 3 dimensional solutions should be specified, though if the samples are complex and especially if there are many samples, it may be worth looking at the 4 dimensional solution.

The goodness of fit of each assessors configuration is measured by the STRESS, where a low value (near zero) implies that the assessors configuration is a good fit, whilst a value of 1 indicates a poor fit. This index decreases as the number of dimensions increase. The $R^2$ value increases as the number of dimensions increase, and is a measure of how well the data fits the model; the higher the value (near 1) the better the fit.
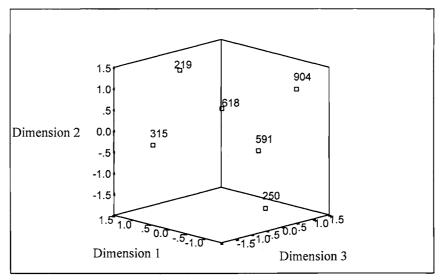
Table 9.4 shows the stress and $R^2$ values for Panel Y, where these values are provided for each assessor and an overall average. Each assessor contributes to the overall STRESS, and it can be seen that Assessors 1 and 7 make the largest contribution. In terms of the $R^2$ values, Assessors 3 has data that corresponds well to the model.

**Table 9.4:** Stress and RSQ results for the 8 assessors in Panel Y.

| Assessor (Matrix) | Stress | RSQ |
|:---:|:---:|:---:|
| 1 | 0.262 | 0.544 |
| 2 | 0.148 | 0.661 |
| 3 | 0.072 | 0.976 |
| 4 | 0.145 | 0.703 |
| 5 | 0.120 | 0.660 |
| 6 | 0.157 | 0.806 |
| 7 | 0.197 | 0.630 |
| 8 | 0.166 | 0.584 |
| Average | 0.167 | 0.700 |

As well as individual sample configurations, INDSCAL provides a consensus, which is shown as a three dimensional representation in Figure 9.1

Figure 9.1: Derived stimulus configuration using the individual differences (weighted) Euclidean distance model.

The samples appear to be separated on all three dimensions, though whether this is significant or not needs further interpretation.

The next output is the assessor (subject) weights, which measure the importance of each dimension to each assessor. These are shown in Table 9.5, and reveal that Assessor 3 uses only one dimension, whilst the other assessors use 2 or 3 dimensions.

**Table 9.5**: Assessor weights on 3 dimensions together with the corresponding weirdness measure.

| | | Dimension | | |
|---|---|---|---|---|
| **Assessor (Matrix)** | **Weirdness** | **1** | **2** | **3** |
| 1 | 0.514 | 0.726 | 0.123 | 0.047 |
| 2 | 0.637 | 0.496 | 0.625 | 0.154 |
| 3 | 0.814 | 0.986 | 0.043 | 0.032 |
| 4 | 0.019 | 0.770 | 0.248 | 0.222 |
| 5 | 0.418 | 0.559 | 0.333 | 0.487 |
| 6 | 0.432 | 0.881 | 0.086 | 0.152 |
| 7 | 0.528 | 0.703 | 0.081 | 0.359 |
| 8 | 0.084 | 0.691 | 0.255 | 0.202 |
| Average | | 0.550 | 0.082 | 0.063 |

Table 9.5 also shows the weirdness associated with each assessor. An assessor with weights proportional to the average weights has a weirdness of zero (Assessors 4 and 8), whilst an assessor with one large weight and many small weights has a weirdness near 1 (Assessor 3).

# 10.  RV COEFFICIENT

## 10.1 Principle

A simple way to measure whether the attribute information from the profile, provided by an assessor or a panel, corresponds to the expected result is to calculate the correlation coefficient between the panel data and the consensus result, for each common attribute. The correlations measure the strength of the relationship between the observed data and the expected (consensus) result.  If the correlation is one, then the panel has performed perfectly.  A correlation of zero would imply no relationship; the panel did not perform well, or there were no perceivable differences between any of the samples.  The latter should not be the case for proficiency testing scheme samples.  A negative correlation would suggest the scoring was done in reverse.

However, the simple correlation coefficient is a univariate measure, whereas sensory profile data from any panel comprises data from different vocabularies.  A more powerful tool would be to measure the multivariate correlation between sample configurations obtained from a suitable multivariate mapping method.  Such a method can be found in the RV coefficient (Robert and Escoufier, 1976), and measures the similarity between two configurations of p dimensions and n observations (samples).  The RV coefficient is sometimes referred to as the 'index of association' between two configurations.

## 10.2 Procedure

The RV coefficient is used to measure the strength of the association between two matrices, with the same number of rows, and in this application the same number of samples*replicates.  For example, Table 10.1 shows the scores for the sample map for Assessor 1 in Panel Y and the corresponding scores for the consensus sample map.

**Table 10.1:** Sample scores on the first two GPA dimensions for Assessor 1 and the consensus configurations.

| | Consensus | | Assessor 1 | |
|---|---|---|---|---|
| **Product** | **Dim 1** | **Dim 2** | **Dim 1** | **Dim 2** |
| 219 | -0.44 | -0.08 | 0.473 | 0.142 |
| 219 | -0.46 | -0.28 | -0.405 | -0.077 |
| 219 | -0.90 | 0.32 | -0.665 | -0.091 |
| 250 | 0.55 | 0.40 | 0.452 | 0.265 |
| 250 | 1.00 | -0.08 | 0.568 | -0.110 |
| 250 | 0.81 | -0.15 | 0.714 | -0.006 |
| 315 | -0.74 | 0.63 | -0.390 | 0.149 |
| 315 | -0.43 | -0.22 | -0.383 | 0.015 |
| 315 | -0.40 | -0.41 | -0.692 | -0.208 |
| 591 | -0.96 | 0.31 | -0.855 | -0.090 |
| 591 | -0.69 | -0.17 | -1.000 | 0.003 |
| 591 | -0.65 | -0.58 | -0.987 | -0.241 |
| 618 | 0.65 | 0.20 | 0.377 | 0.168 |
| 618 | 0.56 | -0.38 | 0.666 | -0.105 |
| 618 | 0.64 | 0.35 | 0.661 | 0.140 |
| 904 | 0.43 | 0.62 | 0.373 | 0.271 |
| 904 | 0.61 | -0.35 | 0.438 | -0.144 |
| 904 | 0.45 | -0.16 | 0.654 | -0.081 |

The RV takes the range from 0 to 1, with 1 indicating a perfect association, and 0 indicating no association between two matrices, or configurations. There is no statistical test of significance for the RV value, but according to Schlich and Guichard (1989), a value of 0.95 indicates a good similarity level. This is based on the fact that RV is analogous to the squared correlation coefficient.

In the case of the sample maps derived from GPA for Panel Y, the RV coefficient can be measured between the 2 dimensional sample map for each assessor and that of the consensus to give the results shown in Table 10.2. It can be seen that the assessors are not in good agreement with the consensus, which indicates that there were differences between the assessors.

**Table 10.2**: RV coefficient between the 2 dimensional sample maps for each assessor in Panel Y and the consensus map, as derived from GPA.

| Assessor | RV |
|----------|------|
| 1 | 0.74 |
| 2 | 0.49 |
| 3 | 0.88 |
| 4 | 0.57 |
| 5 | 0.55 |
| 6 | 0.70 |
| 7 | 0.58 |
| 8 | 0.63 |
| Average | 0.64 |

# 11. SUMMARY OF PERFORMANCE FOR 1999 RING TRIAL

## 11.1 Means, Analysis of Variance and Multiple Comparisons

**Data Analysed**

To look at the univariate methods of analysis, only the 5 common attributes (overall odour, overall flavour, sweet, salty and thickness) were considered in detail. Further results with respect to the other attributes are provided in the Appendices.

In this section, the data for each panel were analysed separately, but in order to establish an 'expected result', a global analysis was also conducted.

**Table of Means**

Appendix 2 shows the sample means on the 5 common attributes, measured in the original scale used by the panel (Table 9.3).

**Analysis of Variance**

Tables 11.1 to 11.5 show the results of analysis of variance on the 5 common attributes for each panel. In the first analysis, sample was specified as a main effect, whilst a second analysis specified cornflour and flavour as the main effects (with interaction). In both cases, assessors were considered as random effect.

Considering attributes discriminating between samples at the 5% significance level, only Panels U and Z show a significant sample effect for overall odour. However, when the analysis is broken down into the sample ingredients, Panels P, R, U and Z could distinguish between samples with and without added cornflour. No panel, with the

exception of Panel Y, discriminated (at the 5% level of significance) between the 2 levels of flavour, and none of the panels had a significant interaction between cornflour and flavour.

In terms of overall flavour, Panels U, Y and Z were able to discriminate between samples, but Panels R, T, U, V, Y and Z were able to distinguish between samples with and without added cornflour. However, only Panel Y used overall flavour to discriminate between different levels of added flavour.

For sweetness, Panels N, T, U and Y could discriminate between the samples, and in particular discriminated between samples with or without added cornflour. However, only Panel T could detect differences in sweetness with respect to added flavour.

With respect to salty, Panel U discriminated between the samples, and those with and without added cornflour. Panel Z also discriminated between the samples, but this was on the basis of added flavour.

All panels could differentiate the samples in terms of thickness, and this was mainly with respect to differentiating samples with and without added cornflour.


**Multiple Comparisons**

It is possible to undertake a multiple comparison test on the common attributes, using Tukey's HSD (Honestly Significant Difference). This was not undertaken here, but the method is demonstrated for the sensory dimensions from the GPA (Section 11.3).

**Table 11.1**: Analysis of variance on *overall odour* for each panel: p-values for sample effect, and for effect of cornflour and flavour.

| | | Ingredient ANOVA | | |
|---|---|---|---|---|
| **Panel** | **Sample** | **Cornflr** | **Flavour** | **C*F** |
| N | 0.782 | 0.531 | 0.958 | 0.397 |
| P | 0.015 | 0.005 | 0.658 | 0.246 |
| Q | 0.285 | 0.243 | 0.053 | 0.458 |
| R | 0.135 | 0.007 | 0.858 | 0.297 |
| T | 0.788 | 0.320 | 0.674 | 0.638 |
| U | 0.000 | 0.000 | 0.357 | 0.043 |
| V | 0.484 | 0.710 | 0.155 | 0.453 |
| W | 0.629 | 0.448 | 0.991 | 0.167 |
| Y | 0.622 | 0.077 | 0.876 | 0.963 |
| Z | 0.014 | 0.002 | 0.192 | 0.809 |

**Table 11.2**: Analysis of variance on *overall flavour* for each panel: p-values for sample effect, and for effect of cornflour and flavour.

| | | Ingredient ANOVA | | |
|---|---|---|---|---|
| **Panel** | **Sample** | **Cornflr** | **Flavour** | **C*F** |
| N | 0.836 | 0.490 | 0.726 | 0.648 |
| P | 0.489 | 0.244 | 0.110 | 0.356 |
| Q | 0.890 | 0.927 | 0.547 | 0.674 |
| R | 0.110 | 0.020 | 0.252 | 0.256 |
| T | 0.124 | 0.013 | 0.841 | 0.274 |
| U | 0.000 | 0.000 | 0.522 | 0.353 |
| V | 0.137 | 0.036 | 0.416 | 0.373 |
| W | 0.518 | 0.455 | 0.243 | 0.228 |
| Y | 0.004 | 0.003 | 0.033 | 0.095 |
| Z | 0.036 | 0.001 | 0.430 | 0.796 |

**Table 11.3**:    Analysis of variance on *sweet* for each panel: p-values for sample effect, and for effect of cornflour and flavour.

| Panel | Sample | Ingredient ANOVA | | |
|:---:|:---:|:---:|:---:|:---:|
| | | Cornflr | Flavour | C*F |
| N | 0.031 | 0.000 | 0.952 | 0.863 |
| P | 0.175 | 0.756 | 0.087 | 0.297 |
| Q | 0.170 | 0.619 | 0.026 | 0.523 |
| R | 0.841 | 0.235 | 0.671 | 0.632 |
| T | 0.007 | 0.030 | 0.004 | 0.189 |
| U | 0.000 | 0.000 | 0.079 | 0.659 |
| V | 0.533 | 0.338 | 0.450 | 0.684 |
| W | 0.508 | 0.451 | 0.715 | 0.352 |
| Y | 0.006 | 0.002 | 0.098 | 0.786 |
| Z | 0.534 | 0.280 | 0.478 | 0.850 |

**Table 11.4**:    Analysis of variance on *salty* for each panel: p-values for sample effect, and for effect of cornflour and flavour.

| Panel | Sample | Ingredient ANOVA | | |
|:---:|:---:|:---:|:---:|:---:|
| | | Cornflr | Flavour | C*F |
| N | 0.373 | 0.301 | 0.389 | 0.336 |
| P | 0.939 | 0.494 | 0.998 | 0.613 |
| Q | 0.309 | 0.354 | 0.139 | 0.660 |
| R | 0.127 | 0.259 | 0.202 | 0.374 |
| T | 0.241 | 0.169 | 0.076 | 0.339 |
| U | 0.000 | 0.000 | 0.661 | 0.390 |
| V | 0.394 | 0.089 | 0.384 | 0.173 |
| W | 0.507 | 0.265 | 0.603 | 0.603 |
| Y | 0.403 | 0.138 | 0.429 | 0.523 |
| Z | 0.064 | 0.812 | 0.013 | 0.237 |

**Table 11.5:** Analysis of variance on *thickness* for each panel: p-values for sample effect, and for effect of cornflour and flavour.

| Panel | Sample | Ingredient ANOVA | | |
|:---:|:---:|:---:|:---:|:---:|
| | | **Cornflr** | **Flavour** | **C*F** |
| N | 0.000 | 0.000 | 0.063 | 0.493 |
| P | 0.000 | 0.000 | 0.465 | 0.116 |
| Q | 0.002 | 0.000 | 0.416 | 0.486 |
| R | 0.000 | 0.000 | 0.201 | 0.005 |
| T | 0.000 | 0.000 | 0.954 | 0.304 |
| U | 0.000 | 0.000 | 0.717 | 0.001 |
| V | 0.000 | 0.000 | 0.702 | 0.873 |
| W | 0.000 | 0.000 | 0.000 | 0.003 |
| Y | 0.000 | 0.000 | 0.178 | 0.628 |
| Z | 0.000 | 0.000 | 0.517 | 0.548 |

## 11.2 Principal Component Analysis

**Data Analysed**

For each panel separately, a principal component analysis (covariance matrix) was undertaken on the data matrix of 18 rows (6 samples x 3 replicates) by p columns, where p was the number of attributes for the panel. Four principal components were specified, and the % variance associated with each for the 10 panels is shown in Appendix 4.

### Analysis of Variance on Consensus Dimensions

Analysis of variance was undertaken on the four principal components; first a one-way ANOVA to look at the overall effect of samples, then a second two-way ANOVA with interaction to look at the effect of the two ingredients. As only the first 2 PCs yield significant results, these are reported in Table 11.6.

The interaction between cornflour and flavour is not shown, as no significant results were found.

**Table 11.6:** Analysis of variance on PCA dimensions for each panel: p-values for sample effect, and for effect of cornflour and flavour.

| | Analysis 1 | | Analysis 2 | | | |
| | Sample | | Cornflour | | Flavour | |
| Panel | PC 1 | PC 2 | PC 1 | PC 2 | PC 1 | PC 2 |
|---|---|---|---|---|---|---|
| N | 0.000 | 0.161 | 0.000 | 0.586 | 0.437 | 0.170 |
| P | 0.000 | 0.582 | 0.000 | 0.724 | 0.737 | 0.198 |
| Q | 0.209 | 0.026 | 0.018 | 0.013 | 0.996 | 0.033 |
| R | 0.000 | 0.704 | 0.000 | 0.779 | 0.742 | 0.430 |
| T | 0.000 | 0.497 | 0.000 | 0.839 | 0.965 | 0.742 |
| U | 0.000 | 0.919 | 0.000 | 0.747 | 0.969 | 0.648 |
| V | 0.000 | 0.891 | 0.000 | 0.868 | 0.974 | 0.519 |
| W | 0.007 | 0.621 | 0.001 | 0.269 | 0.137 | 0.463 |
| Y | 0.000 | 0.006 | 0.000 | 0.820 | 0.238 | 0.001 |
| Z | 0.000 | 0.146 | 0.000 | 0.591 | 0.440 | 0.029 |
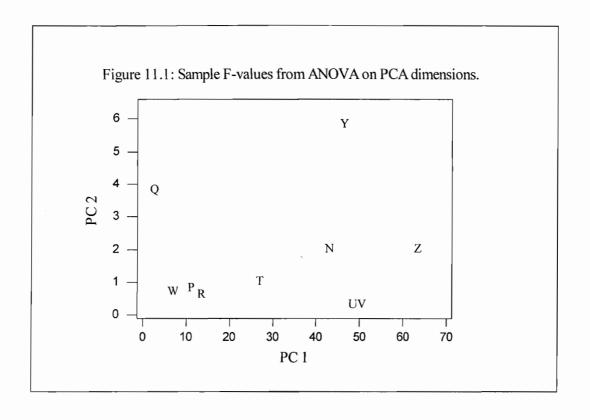
Table 11.7 provides a rough guide to the overall results, though these can be further explored by an F-value plot from the one way ANOVA on PC 1 and PC 2 (Figure 11.1).

**Table 11.7:** Summary of results from principal component analysis.

| Panel | Cornflour | Flavour | General |
|-------|-----------|---------|---------|
| N | PC 1 okay. | Did not discriminate. | |
| P | PC 1 okay. | Did not discriminate. | Differences between replications. |
| Q | PC 1 and PC 2 okay. | | Large differences between replications. |
| R | PC 1 okay. | Did not discriminate. | Differences between replications. |
| T | PC 1 okay. | Did not discriminate. | |
| U | PC 1 okay. | Did not discriminate. | |
| V | PC 1 okay. | Did not discriminate. | |
| W | PC 1 okay. | Did not discriminate. | Differences between replications. |
| Y | PC 1 okay. | Okay PC 2 – none and low vs. high. | |
| Z | PC 1 okay. | Okay PC 2 – none and low vs. high. | |

This graph (Figure 11.1) indicates that Panel Y has discriminated samples on both dimensions, whilst Panels Z, U, V and N were performed well in respect of PC 1. Panel T was average, whilst Panels P, R and W performed poorly in terms of sample discrimination. Panel Q could not discriminate between the samples on PC 1, but was better than most other panels on PC 2.

It is also possible to undertake a multiple comparison test on the sample mean scores on the PC dimensions, but this is summarised for the GPA results.

Figure 11.1: Sample F-values from ANOVA on PCA dimensions.



## 11.3 Generalised Procrustes Analysis – Individual Panels

**Data Analysed**

The data for each panel were analysed by GPA. The data for each panel comprised the sample (6 samples x 3 replicates) by attribute matrix for each assessor. From these analyses, the consensus sample configuration was obtained, as was the sample configuration for each individual assessor.

## Analysis of Variance on Consensus Dimensions
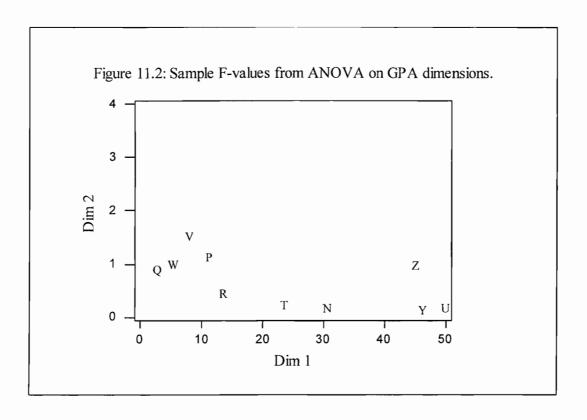
Analysis of variance on the GPA dimensions generally reflected the results from PCA, though the second dimension did not discriminate between the samples. The interaction between cornflour and flavour is not shown, as no significant results were found.

Table 11.8: Analysis of variance on GPA dimensions for each panel: p-values for sample effect, and for effect of cornflour and flavour.

| Panel | Sample | | Cornflour | | Flavour | |
|---|---|---|---|---|---|---|
| | Dim 1 | Dim 2 | Dim 1 | Dim 2 | Dim 1 | Dim 2 |
| N | 0.000 | 0.975 | 0.000 | 0.730 | 0.722 | 0.794 |
| P | 0.000 | 0.404 | 0.000 | 0.417 | 0.972 | 0.253 |
| Q | 0.149 | 0.530 | 0.016 | 0.312 | 0.806 | 0.717 |
| R | 0.000 | 0.828 | 0.000 | 0.781 | 0.803 | 0.552 |
| T | 0.000 | 0.957 | 0.000 | 0.474 | 0.929 | 0.874 |
| U | 0.000 | 0.981 | 0.000 | 0.820 | 0.936 | 0.755 |
| V | 0.002 | 0.265 | 0.072 | 0.299 | 0.001 | 0.670 |
| W | 0.017 | 0.471 | 0.001 | 0.164 | 0.203 | 0.364 |
| Y | 0.000 | 0.991 | 0.000 | 0.618 | 0.072 | 0.933 |
| Z | 0.000 | 0.942 | 0.000 | 0.611 | 0.559 | 0.172 |

Figure 11.2 shows the F-value plot from the one way ANOVA on Dim 1 and Dim 2. It can be observed that Panels U, Y and Z have high discrimination on the first dimension, whereas Panels Q and W, for example, do not discriminate well between the samples. Discrimination on the second dimension is not good, though Panels V, P and Z, for example, perform better than Panels N, T, U and Y.

Figure 11.2: Sample F-values from ANOVA on GPA dimensions.

## Multiple Comparisons on Sample

As only the first dimension on cornflour was used to discriminate between the samples, the multiple comparison results will be restricted to this aspect.

Table 11.9 shows the mean value for the two levels of cornflour on the first GPA dimension, together with the Tukey multiple comparison value. With the exception of Panel V, there were clear differences between the two levels of cornflour.

**Table 11.9:**   Mean value associated with the 2 levels of cornflour, together with the Tukey multiple comparison value at the 5% significance level.

| Panel | Flour Level | | Tukey |
| :---: | :---: | :---: | :---: |
| | 0 | 1 | |
| N | -0.601 | 0.601 | 0.2 |
| P | 0.532 | -0.534 | 0.3 |
| Q | 0.279 | -0.278 | 0.4 |
| R | -0.577 | 0.574 | 0.3 |
| T | -0.643 | 0.644 | 0.3 |
| U | 0.731 | -0.732 | 0.2 |
| V | -0.160 | 0.160 | ns |
| W | -0.398 | 0.397 | 0.4 |
| Y | 0.633 | -0.630 | 0.2 |
| Z | -0.697 | 0.696 | 0.2 |

**RV Coefficient between Each Assessor and the Consensus Sample Map**

Table 11.10 shows the results of calculating the RV coefficient between each assessor's sample configuration and the consensus, for each of the 10 panels.

Table 11.10:  RV coefficient between each assessor's sample map and the consensus map.

| Assessor | Panel | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | N | P | Q | R | T | U | V | W | Y | Z |
| 1 | 0.87 | 0.56 | 0.55 | 0.86 | 0.75 | 0.88 | 0.51 | 0.76 | 0.74 | 0.43 |
| 2 | 0.82 | 0.77 | 0.55 | 0.87 | 0.66 | 0.88 | 0.81 | 0.53 | 0.49 | 0.78 |
| 3 | 0.65 | 0.70 | 0.58 | 0.91 | 0.48 | 0.95 | 0.75 | 0.52 | 0.88 | 0.68 |
| 4 | 0.63 | 0.72 | 0.68 | 0.92 | 0.67 | 0.78 | 0.89 | 0.60 | 0.57 | 0.66 |
| 5 | 0.56 | 0.70 | 0.45 | 0.94 | 0.76 | 0.85 | 0.89 | 0.65 | 0.55 | 0.71 |
| 6 | 0.69 | 0.78 | 0.59 | 0.77 | 0.83 | 0.94 | 0.70 | 0.68 | 0.70 | 0.49 |
| 7 | 0.83 | 0.72 | 0.66 | 0.83 | 0.88 | 0.72 | 0.80 | 0.77 | 0.58 | 0.70 |
| 8 | 0.77 | 0.62 | -- | 0.87 | 0.79 | 0.75 | 0.87 | 0.73 | 0.63 | 0.79 |
| 9 | 0.84 | 0.54 | -- | -- | 0.74 | 0.84 | -- | 0.69 | -- | 0.69 |
| 10 | 0.68 | -- | -- | -- | 0.66 | 0.74 | -- | 0.71 | -- | -- |
| 11 | 0.50 | -- | -- | -- | 0.74 | 0.91 | -- | -- | -- | -- |
| 12 | 0.82 | -- | -- | -- | 0.72 | -- | -- | -- | -- | -- |
| 13 | 0.76 | -- | -- | -- | 0.83 | -- | -- | -- | -- | -- |
| 14 | 0.87 | -- | -- | -- | 0.70 | -- | -- | -- | -- | -- |
| 15 | 0.78 | -- | -- | -- | 0.85 | -- | -- | -- | -- | -- |
| 16 | -- | -- | -- | -- | 0.81 | -- | -- | -- | -- | -- |
| Average | 0.74 | 0.68 | 0.58 | 0.87 | 0.74 | 0.84 | 0.78 | 0.66 | 0.64 | 0.66 |

## 11.4 Generalised Procrustes Analysis – Average Data, All Attributes

**Data Analysed**

The data for each panel were averaged across assessors and replicates.  The data for each panel comprised the sample (6 samples) by attribute matrix.  From these analyses, the consensus sample configuration was obtained, as was the sample configuration for each individual panel.

## RV Coefficient between Each Panel and the Consensus Sample Map

From the GPA, the sample map for each panel was compared with the consensus map for similarity using the RV coefficient, and the results of this are shown in Appendix 5. Of more interest is the RV coefficient between the sample map for each panel and the overall consensus (Table 11.11).

**Table 11.11:** RV coefficient between panel's sample map and the consensus, based on all attributes.

| Panel | RV Coefficient |
|---|---|
| N | 0.97 |
| P | 0.97 |
| Q | 0.85 |
| R | 0.98 |
| T | 0.99 |
| U | 0.99 |
| V | 0.99 |
| W | 0.80 |
| Y | 0.98 |
| Z | 0.98 |
| Average | 0.95 |

From these results, it is clear that Panels N, P, R, T, U, V, Y and Z were all above average. However, these results do not completely tie up with those in Table 11.10, and therefore some further investigations are required before considering the approach to take for the second ring trial. However, it is clear that Panels Q and W performed less well than the other panels.

## 11.5 Generalised Procrustes Analysis – Common Attributes

**Data Analysed**

The data for each panel were averaged across assessors and replicates. The data for each panel comprised the sample (6 samples) by common attribute (5 attributes) matrix. From these analyses, the consensus sample configuration was obtained, as was the sample configuration for each individual panel.

**RV Coefficient between Each Panel and the Consensus Sample Map**

From the GPA, the sample map for each panel was compared with the consensus sample map for similarity using the RV coefficient, and the results of this are shown in Appendix 5. Of more interest is the RV coefficient between the sample map for each panel and the overall consensus (Table 11.12). Panels Q and W performed less well than the other panels.

**Table 11.12**: RV coefficient between panel's sample map and the consensus, based on the 5 common attributes.

| Panel | RV Coefficient |
|---|---|
| N | 0.97 |
| P | 0.94 |
| Q | 0.83 |
| R | 0.97 |
| T | 0.99 |
| U | 0.98 |
| V | 0.99 |
| W | 0.77 |
| Y | 0.98 |
| Z | 0.99 |
| Average | 0.94 |

## 11.6 INDSCAL – Individual Panels

**Data Analysed**

For each data set (panel), the data for each assessor was averaged over replicates. An inter-sample distance matrix was then calculated for each assessor, and these similarity matrixes were submitted to INDSCAL as described in Chapter 9.

**Results**

Table 11.13 shows the Stress, $R^2$ and assessor weights for the 2 dimensional INDSCAL solution, whilst Table 11.14 reports this for the 3 dimensional solution. The information for each individual assessor can be found in Appendix 4.

**Table 11.13**: Stress, $R^2$ and average subject weights from INDSCAL on each panel: 2 dimensional solution.

| Panel | Stress | $R^2$ | Weights | |
| --- | --- | --- | --- | --- |
| | | | **Dim 1** | **Dim 2** |
| N | 0.205 | 0.841 | 0.779 | 0.062 |
| P | 0.248 | 0.497 | 0.426 | 0.071 |
| Q | 0.312 | 0.326 | 0.189 | 0.138 |
| R | 0.172 | 0.897 | 0.828 | 0.070 |
| T | 0.216 | 0.751 | 0.628 | 0.123 |
| U | 0.140 | 0.920 | 0.717 | 0.203 |
| V | 0.227 | 0.630 | 0.537 | 0.094 |
| W | 0.270 | 0.476 | 0.378 | 0.098 |
| Y | 0.243 | 0.572 | 0.501 | 0.071 |
| Z | 0.287 | 0.539 | 0.414 | 0.125 |

Table 11.13 indicates that Panel Q had the largest Stress and lowest $R^2$, indicating that the consensus configuration is not a good representation of the samples.

**Table 11.14:** Stress, $R^2$ and average subject weights from INDSCAL on each panel: 3 dimensional solution.

| Panel | Stress | $R^2$ | Weights | | |
|-------|--------|-------|---------|---|---|
| | | | **Dim 1** | **Dim 2** | **Dim 3** |
| N | 0.157 | 0.859 | 0.666 | 0.139 | 0.054 |
| P | 0.169 | 0.667 | 0.467 | 0.116 | 0.084 |
| Q | 0.209 | 0.463 | 0.198 | 0.164 | 0.101 |
| R | 0.146 | 0.913 | 0.643 | 0.183 | 0.086 |
| T | 0.138 | 0.810 | 0.386 | 0.317 | 0.107 |
| U | 0.119 | 0.944 | 0.443 | 0.357 | 0.144 |
| V | 0.164 | 0.640 | 0.511 | 0.092 | 0.037 |
| W | 0.179 | 0.635 | 0.379 | 0.154 | 0.101 |
| Y | 0.167 | 0.700 | 0.550 | 0.082 | 0.063 |
| Z | 0.181 | 0.618 | 0.403 | 0.135 | 0.079 |

Table 11.14 reveals that Panel Q has a high Stress and low $R^2$, followed by Panel Z, indicating that the consensus solution is not a good representation of the samples.

## 11.7 INDSCAL – Averaged Panel Data

**Data Analysed**

Prior to this analysis, the data for each panel were converted to a 0 – 100 scale. For each data set (panel), treatment means were calculated for each attribute. An inter-sample distance matrix was then calculated for each panel, and these similarity matrices were submitted to INDSCAL as described in Chapter 9.

**Results**

Table 11.15 shows the Stress, $R^2$, weirdness and assessor weights for the 2 dimensional INDSCAL solution, whilst Table 11.16 reports this for the 3 dimensional solution. The three dimensional graphic is given below this latter table.

**Table 11.15:** Stress, $R^2$ and average subject weights from INDSCAL on each panel: 2 dimensional solution.

| | | | | Weights | |
|---|---|---|---|---|---|
| **Panel** | **Stress** | **$R^2$** | **Weirdness** | **Dim 1** | **Dim 2** |
| N | 0.168 | 0.920 | 0.520 | 0.954 | 0.104 |
| P | 0.260 | 0.358 | 0.769 | 0.332 | 0.498 |
| Q | 0.290 | 0.408 | 0.176 | 0.600 | 0.218 |
| R | 0.184 | 0.858 | 0.593 | 0.922 | 0.084 |
| T | 0.223 | 0.798 | 0.404 | 0.885 | 0.123 |
| U | 0.119 | 0.946 | 0.755 | 0.971 | 0.052 |
| V | 0.200 | 0.834 | 0.127 | 0.891 | 0.200 |
| W | 0.308 | 0.222 | 0.555 | 0.376 | 0.284 |
| Y | 0.325 | 0.274 | 0.115 | 0.497 | 0.164 |
| Z | 0.347 | 0.199 | 0.186 | 0.418 | 0.154 |
| Average | 0.253 | 0.582 | | 0.531 | 0.050 |
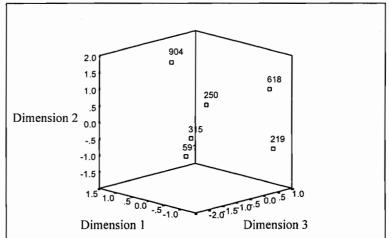
Table 11.15 indicates that Panels Z, Y and W have the largest Stress and low $R^2$, indicating that the consensus configuration is not a good representation of the samples.

**Table 11.16:** Stress, $R^2$ and average subject weights from INDSCAL on each panel: 3 dimensional solution.

| Panel | Stress | $R^2$ | Weirdness | Weights | | |
|---|---|---|---|---|---|---|
| | | | | Dim 1 | Dim 2 | Dim 3 |
| N | 0.119 | 0.911 | 0.453 | 0.945 | 0.098 | 0.089 |
| P | 0.175 | 0.363 | 0.563 | 0.314 | 0.468 | 0.214 |
| Q | 0.220 | 0.342 | 0.180 | 0.529 | 0.168 | 0.184 |
| R | 0.160 | 0.869 | 0.642 | 0.929 | 0.059 | 0.055 |
| T | 0.167 | 0.817 | 0.427 | 0.893 | 0.115 | 0.077 |
| U | 0.099 | 0.964 | 0.855 | 0.981 | 0.014 | 0.032 |
| V | 0.139 | 0.848 | 0.219 | 0.896 | 0.138 | 0.158 |
| W | 0.228 | 0.378 | 0.692 | 0.424 | 0.440 | 0.069 |
| Y | 0.225 | 0.375 | 0.289 | 0.566 | 0.214 | 0.093 |
| Z | 0.074 | 0.936 | 0.679 | 0.691 | 0.158 | 0.658 |
| Average | 0.169 | 0.680 | | 0.568 | 0.056 | 0.057 |

Table 11.16 reveals that Panel W and Y have high Stress and low $R^2$, followed by Panel P and Q, indicating that the consensus solution is not a good representation of the samples.

Figure 11.3: Derived stimulus configuration based on individual differences (weighted) Euclidean distance model.

## Comments Regarding the INDSCAL Procedure

- The SPSS™ (Version 8) program was used for these calculations.
- The methodology is heavily based on the work of Prof. Forrest W Young of the University of North Carolina. Any person who wishes to use this technique, as implemented in SPSS™, for important projects should read the Chapter on Multidimensional Scaling in the technical manual.
- The examples quoted in the text are based on Ratio scaling which is the severest form of scaling.
- Solutions for two dimensions are not nested in three dimensions. Consequently adding further dimensions will not necessarily improve the fit of the underlying model.
- Subsequently we have repeated these calculations using "ordinal" scaling (the ALSCAL procedure). This reduces the "Stress" and increases the $R^2$ statistic, so apparently the model fits the data better.
- The data sets that we used to illustrate the use of this technique are arguably too small and with too simple a structure to benefit from this technique. Nevertheless, it is the only well described technique that can allow for different sensitivities to each sensory dimension.

## Use of INDSCAL for Future Ring Trials

While INDSCAL may offer potential for analysing descriptive profile data for ring trials in a proficiency testing scheme, it is not explored for further data as part of the PROFISENS project.

# 12. PROCEDURE FOR CALCULATING EXPECTED RESULTS

## 12.1 Introduction

This chapter outlines the first attempt at establishing a scheme to evaluate the performance of panels in a proficiency testing scheme for sensory profiling. In proposing this scheme it should be highlighted that the actual criteria are at this stage illustrative, as these will be set by the Proficiency Testing Provider based on the results of screening work undertaken prior to a ring trial. Moreover, much further work is still required to think through the methods proposed.

One key issue that arose from the first ring trial on tomato soup was the validity of comparing panel data against the consensus result. The reasoning behind this was that the panels involved in the ring trial influence the consensus. Consequently, a good panel may be down graded if all other panels are very good.

To get round this problem it is proposed to use pre-testing with 2-4 panels to establish the expected result, and to allocate levels of performance in relation to this. The choice of these panels needs careful consideration.
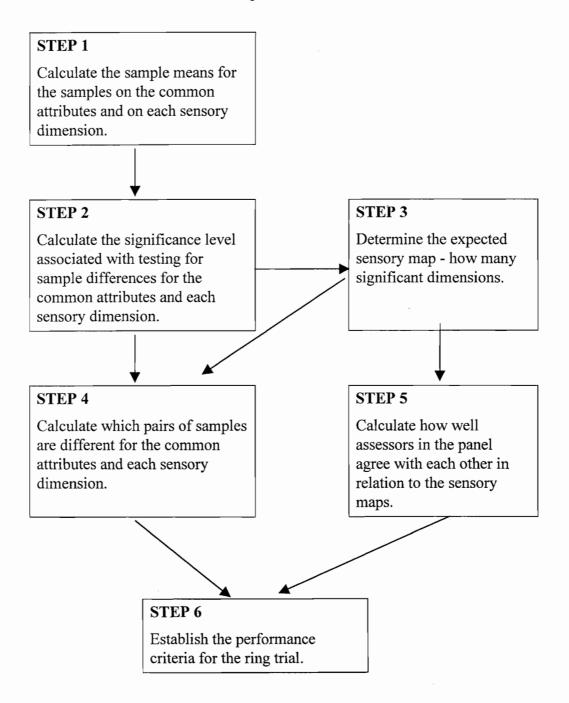
## 12.2 Establishing the Expected Result

### Step 1 – Calculate the Sample Means

For each panel in the pre-test, calculate the sample means for each common attribute and for each sensory dimension (from PCA or GPA). If all pre-test panels are in good agreement, then an 'expected sample order (means)' can be specified for each common attribute and sensory dimension. If there is some disagreement, then Steps 2 and 3 will

help establish if this is because samples were 'switched' in rating because there was no perceptible difference between them. The Pearson correlation is then determined between the 'expected sample means' and the actual panel sample means at the 10% level of significance. This level of significance is chosen to eliminate the possibility of downgrading a panel because two or more samples were not perceptibly different.

*Possible performance criteria*:    score 0 if $p > 0.10$    or    negative correlation
                                         score 1 if $p \leq 0.10$

---

**STEP 1**

Calculate the sample means for the samples on the common attributes and on each sensory dimension.

↓

**STEP 2**

Calculate the significance level associated with testing for sample differences for the common attributes and each sensory dimension.

**STEP 3**

Determine the expected sensory map - how many significant dimensions.

**STEP 4**

Calculate which pairs of samples are different for the common attributes and each sensory dimension.

**STEP 5**

Calculate how well assessors in the panel agree with each other in relation to the sensory maps.

**STEP 6**

Establish the performance criteria for the ring trial.

---

**Step 2 – Calculate the significance level associated with testing for sample differences**

To establish how well each panel discriminated between the samples, analysis of variance should be undertaken on each common attribute and on each sensory dimension. For the common attributes a two-way analysis of variance with interaction between samples and assessors should be used, where assessors are a random effect. If all panels performed well (i.e. $p \leq 0.01$ (1% significance)) on all attributes, then Step 3 may be required to determine if the test was too easy, in other words the panel was able to discriminate between most of the samples in the profile.

In order to establish discrimination ability for the profile as a whole, either principal component analysis (PCA) or generalised Procrustes analysis (GPA) should be undertaken on the data averaged over assessors (replication is retained). A one-way analysis of variance specifying the sample as the main effect should then be undertaken, and the number of dimensions significant at $p \leq 0.05$ (5% significance) retained. If all panels performed well (i.e. $p \leq 0.01$ (1% significance)) on all dimensions, then Step 3 may be required to determine if the test was too easy, in other words the panel was able to discriminate between most of the samples in the profile.

Before deciding the 'expected significance level' for each common attribute and sensory dimension, there should be confidence that the decisions based on the pre-test results will allow some panels in the main test to perform better than the expected result. At the same time the criteria should still allow panels who perform worse than the expected result to be detected.

*Possible performance criteria*: For each attribute and each sensory dimension, the following scoring system could be proposed.

    Score 0 if $p > 0.05$
    Score 1 if $p \leq 0.05$
    Score 2 if $p \leq 0.01$
    Score 3 if $p \leq 0.001$

The scores could then be totalled to provide an overall score for Step 2.

## Step 3 – Determine the expected sensory map

If the panel has performed well, then it would generally be expected that they have a larger number of significant multivariate dimensions than a panel who performed poorly. Given that the 5% significance level is set, then the following scheme may be used.

*Possible performance criteria:*

> Score 0 if no significant dimensions
> Score 1 if 1 significant dimension
> Score 2 if 2 significant dimensions
> Score 3 if 3 significant dimensions

## Step 4 – Differences between samples

Having established an expected significance level for each common attribute and sensory dimension, the next step is to determine which pairs of samples are different at a specified level of significance (for example, 1%, 5% and 10% significance). This can be achieved through the use of a suitable multiple comparison test, for example Tukey's HSD method. From these results the 'expected sample differences' can be set for each common attribute and sensory dimension.

*Possible performance criteria over all significant dimensions (5%):*

> Score 0 if 0 or 1 significant differences
> Score 1 if 2 - 4 significant differences
> Score 2 if 5 - 7 significant differences
> Score 3 if 8 - 10 significant differences
> Score 4 if 11 - 12 significant differences
> Score 5 if 13 - 14 significant differences

*Possible performance criteria for each sensory attribute:*

> Score 0 if 0 or 1 significant differences
> Score 1 if 2 - 4 significant differences
> Score 2 if 5 - 7 significant differences
> Score 3 if 8 - 10 significant differences
> Score 4 if 11 - 12 significant differences
> Score 5 if 13 - 14 significant differences

**Step 5 – Agreement between assessors**

A GPA should be undertaken on each panel's data, and a sample map obtained for each assessor in the panel. The RV coefficient is then calculated between each assessor and the results averaged (RV1), and between each assessor and the panel consensus (RV2), and the results averaged. An RV of '1' indicates perfect agreement, whilst an RV of '0' illustrates no agreement.
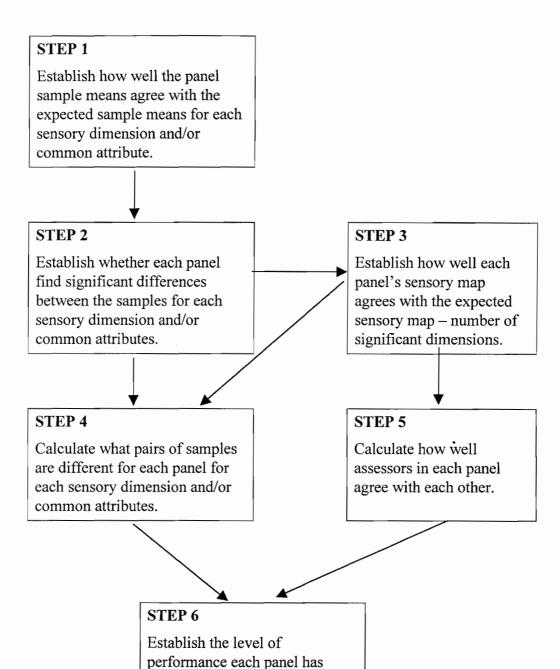
*Possible performance criteria RV1:*   0 if RV < 0.50
                                        1 if RV ≥ 0.50
                                        2 if RV ≥ 0.60
                                        3 if RV ≥ 0.70
                                        4 if RV ≥ 0.80
                                        5 if RV ≥ 0.90

*Possible performance criteria RV2:*   0 if RV < 0.50
                                        1 if RV ≥ 0.50
                                        2 if RV ≥ 0.60
                                        3 if RV ≥ 0.70
                                        4 if RV ≥ 0.80
                                        5 if RV ≥ 0.90

**Step 6 – Establish the performance criteria**

Step 6 involves adding the scores from Steps 1-5 together, and allocating a performance level for the intervals of score. An example of this is not provided, as further work is required in this area.

## 12.3 Determining the Actual Panel Performance

The diagram overleaf illustrates the steps required to analyse each panel's data, to obtain their performance score for each Step, and the overall score.

**STEP 1**

Establish how well the panel sample means agree with the expected sample means for each sensory dimension and/or common attribute.

**STEP 2**

Establish whether each panel find significant differences between the samples for each sensory dimension and/or common attributes.

**STEP 3**

Establish how well each panel's sensory map agrees with the expected sensory map – number of significant dimensions.

**STEP 4**

Calculate what pairs of samples are different for each panel for each sensory dimension and/or common attributes.

**STEP 5**

Calculate how well assessors in each panel agree with each other.

**STEP 6**

Establish the level of performance each panel has achieved.

## 12.4 Modifications for Structured Samples

The illustration for determining the 'expected result' has concentrated on the case where the samples have no structure, normally because they have been selected from the market place. However, as in the case of Tomato Soup, ingredients were modified according to an experimental design. In this case it may be expected that certain differences will be detected between different levels of the ingredients. Therefore, the protocol described in Section 12.2 would be modified to take additional performance criteria into account.

## 12.5 Some Issues to Consider

One important issue to consider during ring trials is whether it is realistic to expect different sensory panels to use the same number of sensory dimensions to describe the key differences between a set of samples. This will be explored further in a future report (McEwan, 2001).

# REFERENCES

Arnold, G.M. and Williams, A.A. (1986). The Use of Generalised Procrustes Techniques in Sensory Analysis. In: Piggott, J.R. (Ed.). Statistical Procedures for Food Research, pp233-253. London: Elsevier Applied Science Publishers.

Gower, J.C. (1975). Generalized Procrustes Analysis. Psychometrica, 40(1), 33-51.

Hunter E.A. (1996). Experimental design. In: Næs, T. and Risvik, E. (Eds.). Multivariate Analysis of Data in Sensory Science, p. 37-69. Amsterdam: Elsevier.

Jones, B. and Kenward, M.G. (1989). Design and Analysis of Cross-over Trials. London: Chapman and Hall.

Jones,B. and Wang, J. (2000). The analysis of repeated measurements in sensory and consumer studies. Food Quality and Preference 11, 35-41.

Lea, P., Næs, T. and Rødbotten, M. (1997). Analysis of Variance for Sensory Data. Chichester: John Wiley and Sons.

Lyon, D.H. (2001). Guidelines for Proficiency Testing in Sensory Analysis. CCFRA. In preparation.

MacFie, H.J.H., Greenhoff, K., Bratchell, N. and Vallis, L. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. Journal of Sensory Studies 4, 129-148.

McEwan, J.A. (2000). Proficiency Testing for Sensory Ranking Tests: Statistical Guidelines. Part 1. R&D Report No. 118. CCFRA.

McEwan, J.A. (2001). Proficiency Testing for Sensory Profiling Tests: Statistical Guidelines. Part 2. R&D Report. CCFRA. In preparation.

Muir, D.D. and Hunter, E.A. (1991/2). Sensory Evaluation of Cheddar Cheese: Order of Tasting and Carryover Effects. Food Quality and Preference **3**, 141-145.

O'Mahony, M. (1986). Sensory Evaluation of Food: Statistical Methods and Procedures. New York: Marcel Dekker, Inc.

Robert, P. and Escoufier, Y. (1976). A Unifying Tool for Linear Multivariate Statistical Methods: the RV Coefficient. Applied Statistics, **25**, 257-265.

Schlich, P. and Guichard, E. (1989). Selection and Classification of Volatile Compounds of Apricot Using the RV Coefficient. Journal of Agricultural and Food Chemistry, **37**, 142-150.

Williams, E.J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. Australian Journal of Scientific Research, Series A, **2**, 149-168.

Young, F.W. and Hamer, R.M. (1987). Multidimensional Scaling: History, Theory and Applications. New Jersey: Lawrence Erlbaum Associates.

# TECHNICAL APPENDIX:
# LINEAR MODELLING OF UNIVARIATE DATA

## Introduction

The best known technique for the analysis of sensory profile data, one descriptor at a time, is Analysis of Variance (ANOVA). Simple versions of this technique are available in widely available software e.g. Excel™, Minitab™ and SPSS™. More sophisticated implementations of this technique are available in Genstat™ and SAS™. In its simplest form, Analysis of Variance assumes that the data are in a complete factorial arrangement of Assessors, Samples and Replicates i.e. there is one unit of data for each combination of the factors. A properly designed sensory experiment usually has this property. The advantage of the complete data matrix is that the variation between experimental units can be divided into a number of additive components which are not dependent on order of fitting. Variance ratio (F) tests of statistical significance can then easily be made. Due to the balanced design, the treatment estimates are found from arithmetic means. Estimates of the variability of the treatment estimates (e.g. standard errors) are obtained from the Residual mean square in the ANOVA. However, simple versions of ANOVA are intolerant of missing data arising.

For sensory data "Order" effects are known to be important (Muir and Hunter, 1991/2). A Sample tested first in a Session is usually rated differently from the same Sample tested later in a Session. Such is the magnitude of this effect that it is important to either randomise Order of presentation within a Session or alternatively design it into the experiment. Experience suggests that this can be effectively done using Latin Square designs. Furthermore, if the cyclic Latin Squares due to Williams (1949) are used as a base for the design then protection is provided against interference effects from the Previous Sample (MacFie *et al.*, 1989; Hunter, 1996). Although there is very little evidence to show that these interference effects are important (Muir and Hunter, 1991/2), it is prudent to design the experiment so that Sample estimates are protected. Sensory experiments clearly fall into the cross-over design category and there are interesting analogies with their

design and analysis in both clinical and pre-clinical medical experiments (Jones and Kenward, 1989).

There is provision in most statistical programs to use a more general technique than ANOVA that is less reliant on a complete factorial data array. This technique is called the "general linear model" and is effectively linear regression with factors. It can, for instance, be used to derive Order effects and the effects of Previous Samples which are generally not fully orthogonal to Samples i.e. additive in the analysis of variance. It can also be used for incomplete data. The disadvantage of the general linear model is that simple versions only recognise one component of variation in the model. SAS™ provides a particularly flexible framework for this technique.

Very simple implementations of ANOVA assume that all factors are fixed. However, when analysing sensory profile data, it is usual to regard Assessors as a random factor and the consequence is that Samples mean squares should be tested against Assessors x Samples mean squares. Provided the data matrix is complete, simple implementations of ANOVA will provide the mean squares and the data analyst can then perform variance ratio (F) tests and derive appropriate standard errors for the tables of means. More sophisticated implementations of ANOVA will recognise both fixed and random factors and thus reduce the need for hand computations.

Recent statistical developments allow general mixed models, in which there is more than one component of variance, to be fitted to sensory data. The simplest form of an experiment with two components of variance is the split-plot experiment which is well known to all those that have studied statistics in connection with agriculture. Suitable algorithms for the general mixed model are programmed in Genstat™, SAS™ and BMDP™. The advantage of this technique is that more appropriate models can be fitted to the data than with traditional ANOVA. Thus the data can be thoroughly explored and important features identified. In particular it is possible to cope with incomplete data from Assessors, allow for Order effects and test for effects of Previous Samples. The overall variation can be split into a number of parts and hence better estimates of error derived,

together with more appropriate tests of significance. In properly designed experiments the Samples estimates are only marginally affected by fitting more complicated models. Jones and Wang (2000) provides an account of fitting sophisticated models to sensory data using SAS™.

**Example of Analysis of Variance**

The data selected for detailed study are the "Total Flavour Intensity" data from Lab-Y. Six Samples were profiled by eight Assessors. There were three Replicates in two Sessions of three Samples per Session. A Williams Latin Square design was used so the effects of Order (of presentation) within a Session and effect of Previous Samples were taken into consideration in the design although the design is not fully orthogonal for these factors and for their interactions with Assessors, Samples and Replicates i.e. the sums of squares are not additive in the analysis of variance.

The simplest form of ANOVA is to consider Assessors as blocks and thus to analyse the data as a randomised block experiment (albeit one in which each Samples is assessed three times in each block). The analysis below assumes that both Assessors and Samples are fixed effects. It is used as a basis from which more appropriate models are developed. The results shown were produced using Genstat™ (version 4 for Windows).

```
Analysis of Variance

Source of variation     d.f.        s.s.        m.s.      v.r.   F pr.

Assessor                  7      10941.67     1563.10     53.26
Sample                    5        607.92      121.58      4.14  0.002
Residual                131       3844.42       29.35

Total                   143      15394.00


         Sample     219       250        315        591       618       904
                   53.62     53.37      53.71      53.21     51.04     48.04

         s.e.d.               1.564
```

From this it can be seen that the total variation of 15394 is divided into three parts, 10941 due to Assessors, 607 due to Samples and 3844 remaining. The variance ratio (F) test for

Assessors gives a value of 53.26 which is clearly of importance and 4.14 for Samples which is statistically significant (p=0.002) but an order of magnitude smaller than the effect of Assessors. The results for Assessors can be explained by each of them using different parts of the scale to rate the Samples. It is not usual to be concerned about Assessors effect being significant. Whenever there are large main effects of factors it is prudent to check for interactions. In this case there is sufficient data to do so because the experiment was replicated three times. The Assessors x Samples interaction from the Residual given above can thus be partitioned to obtain a better estimate of the error.

```
Assessor x Sample       35      983.08      28.09    0.94  0.566
Residual                96     2861.33      29.81
```

In this particular case, but not generally, this interaction is very small (F=0.94) and can be safely ignored when testing the statistical significance of treatment effects. The above calculations are done on the assumption that both Assessors and Samples are fixed effects.

An alternative way of looking at the data is to regard Assessors as random effects. This can be justified by regarding Assessors as drawn from a notional population of potential Assessors. The consequences is that the Samples effects must now be tested against the Assessors x Samples interaction. The ANOVA now has two terms that can be used as error terms depending on which effects are being tested.

```
***** Analysis of variance *****
```

| Source of variation | d.f. | s.s. | m.s. | v.r. | F pr. |
|---|---|---|---|---|---|
| Assessor | 7 | 10941.67 | 1563.10 | 55.65 | |
| Sample | 5 | 607.92 | 121.58 | 4.33 | 0.004 |
| Residual(1) | 35 | 983.08 | 28.09 | 0.94 | |
| Residual(2a) | 96 | 2861.33 | 29.81 | | |
| Total | 143 | 15394.00 | | | |

The mean square from Residual (1) (=Assessors x Samples interaction), which is usually larger than the mean square from Residual (2a), is used to determine the standard errors of

differences for the Tables of Means using the usual formulae. A final improvement to this analysis can be made by partitioning Replicate from Residual (2a). This gives:

```
Replicate                     2      13.88      6.94    0.23  0.796
Residual(2b)                 94    2847.46     30.29
```

This allows the Replicate effect to be tested. Should the Samples or the Assessors change over the course of the experiment, evidence in the form of significant Replicate effects would be expected.

Although only the most sophisticated packages will easily allow such a form of analysis of variance, even the most simple will allow the sums of squares to be determined for each factor and for the interactions. Provided Residual (1) is the Assessors x Samples interaction and that Residual (2a) is the sum of all the remaining interactions, a proper ANOVA can be put together from the components.

The Samples term in the analysis assumes that the they do not have structure. In this particular case the Samples have a simple (complete) factorial structure. There are two levels of Cornflour and three levels of Flavour. The factorial combinations define the six Samples. The Analysis of Variance can reflect this structure and the Samples term can be broken down into three additive terms.

```
Cornflour                     1     261.36    261.36    9.31  0.004
Flavour                       2     206.17    103.08    3.67  0.036
Cornflour x Flavour           2     140.39     70.19    2.50  0.097
```

Further Flavour is a quantitative factor and it is supposed that the three levels are at equal intervals. By using this information the Flavour term can be broken into two components reflecting this, likewise the interaction term.

```
Cornflour                     1     261.36    261.36    9.31  0.004
Flavour                       2     206.17    103.08    3.67  0.036
   Lin                        1     204.17    204.17    7.27  0.011
   Quad                       1       2.00      2.00    0.07  0.791
Cornflour x Flavour           2     140.39     70.19    2.50  0.097
   Cornflour x Lin            1     140.17    140.17    4.99  0.032
   Cornflour x Quad           1       0.22      0.22    0.01  0.930
```

It can be seen that within the range of the factor the response to Flavour is almost completely linear. We can thus be reasonably sure that by increasing the Flavour one further increment beyond the current highest we will continue to see a response. There is also an interesting interaction of this linear effect with Cornflour. This indicates that the linear response to Flavour is different at the two levels of Cornflour.

By considering the structure, the Samples term has been partitioned into a number of meaningful additive terms. These allow tests of significance to be performed in the ANOVA table without the dangers of over-testing that are always present when t-testing differences between means in a large table.

Parallel to the partition of the degrees of freedom in the ANOVA, the Tables of Treatment effects can be arranged in a more meaningful way.

```
Cornflour        no       yes
              50.82     53.51

             sed=0.883

 Flavour       none      low      high
              50.63     52.33    53.54

             sed=1.082

Cornflour Flavour      none      low      high
       no              48.04    51.04    53.38
      yes              53.21    53.62    53.71

             sed=1.530
```

Although it has been shown how the ANOVA can be used in a sophisticated ways the model still does not include a term for Order (of presentation) effects because it is not fully "orthogonal" to other factors and hence cannot easily be incorporated in ANOVA. Although the "general linear model" does allow such data to be analysed, intrinsically it only has one component of variance and so only one error estimate. This limits it's utility in analysing sensory data.

## A General Specification of the Mixed Model

The great advantage of using more general specification of the Mixed Model is that it allows models to be fitted to the data that more fully reflect the error structure. By allowing the error structure to be explored more fully, estimates of error are obtained for the fixed effects that lead to more appropriate tests of significance. From fuller knowledge of the variation it is occasionally possible to identify deficiencies in the sensory procedures.

In contrast to ANOVA where often all factors are treated as fixed effects, in the general mixed model there is a very clear distinction between fixed terms that are primarily Sample terms on one hand and random terms on the other. Nuisance parameters such as Replicate and Order may be regarded as fixed terms that may nevertheless be informative of the conduct of the experiment. For example, large Order effects may be an indicator of fatigue or of insufficient care on the Assessors part in preparing for a new Samples. Large Replicate effects could indicate that the Samples are changing over the course of the experiment or possibly that the Assessors are becoming more proficient as they gain experience with the Samples.

In the analysis below, random effects are Assessors effect and the interactions with Samples, Order and Replicate. The first of these interactions is frequently important and may arise because of confusion about the precise meaning of the descriptor. Random effects are parameterised by a variance component that for present purposes is assumed to come from a Normal distribution.

This example continues by running a mixed model with fixed terms Replicate, Order and Cornflour*Flavour (=Samples). The random terms are set to be Assessors and Assessors x Samples interaction. In addition the remaining variation is set to be a third random component, called "*units*" in the Genstat™ notation. The output is given below:

```
*** Estimated Variance Components ***

Random term            Component
Assessor                  83.57
Assessor x Sample          0.00
*units*                   30.17

*** Deviance: -2*Log-Likelihood ***
                Deviance   d.f.
                 650.27    132


*** Wald tests for fixed effects ***
    Fixed term            Wald statistic      d.f.        p
    Replicate                  0.5              2       0.389
    Order                      0.8              2       0.335
    Cornflour                  8.4              1       0.002
    Flavour                    6.5              2       0.019
    Cornflour x Flavour        4.6              2       0.050

        Replicate          1              2              3
                         52.35          51.73          52.42

Standard error of differences:       1.121

        Order              1         2         3
                         52.34     51.97     52.19

Standard error of differences:    Average    1.148


*** Table of predicted means for Cornflour ***

    Cornflour           no          yes
                       50.83        53.51

Standard error of differences:    0.9177


    Flavour            none         low         high
                       50.67        52.30       53.54

Standard error of differences:    Average    1.143


    Flavour            none         low         high
    Cornflour
        no             48.09        51.02       53.37
        yes            53.25        53.57       53.71

Standard error of differences:    Average    1.601
```

The estimates of the components of variance show that there is a large component for Assessors and no statistical evidence for an Assessors x Samples interaction. These estimates are entirely consistent with those from the ANOVA.

Tests of significance are carried out using the Wald statistic which is distributed as a $\chi^2$. This is analogous to the variance ratio (F) test of the ANOVA.

More complicated models of variation can be tested by increasing the number of random terms in the model. Arguably a more complicated model will have components of variance for Assessors x Session and Assessors x Order interactions. The results for the components of variance are:

```
*** Estimated Variance Components ***

Random term              Component
Assessor                     82.85
Assessor x Sample             0.00
Assessor x Session            2.73
Assessor x Order              4.13
*units*                      25.12



*** Deviance: -2*Log-Likelihood ***
                  Deviance   d.f.
                    646.79   130
```

A test of the difference between these two models is performed by differing the Deviances and df and estimating the probability from a $\chi^2$ distribution.

In this particular case, the extra two components in the model reduce the deviance by 3.48 (2 df) which is not significant. Readers should not, however, assume that the more complicated models advocated are seldom required. In too many sets of sensory data there is clear statistical evidence of a component of variance for an Assessors x Samples interaction. This can mean that the Assessors use different amounts of scale, but essentially order the Samples the same way. However, there is a possibility that one or more Assessors rate the Samples differently from the others and this may be because of their interpretation of the attribute or because they are insensitive or hyper-sensitive to the attribute. Allowing for the Assessors x Samples component of variance in our models can radically alter our tests of significance of the Samples effect.

## Results of Fitting the Mixed Model to Panels Y, Z and R

The more complicated model outlined above was fitted to the re-scaled data from panels Y, Z and R. The components of variance and the Wald tests are given in Tables 1, 2 and 3.

The components of variance for Assessors are large relative to the Residual error for Panel Y, smaller for Panel Z and smallest for Panel R. This is consistent with the Assessors of Panel Y using different parts of the scale. Overall there is little evidence of significant components of variance for Assessors x Samples, Assessors x Session and Assessors x Order. However, readers should not assume that such components are invariably small and of no interest. The high Residual error for "Thickness" (Panel Y) and for Mouthfeel and Texture (Panel R) is evidence for confusion about these attributes.

Fixed effects are tested using the Wald test. The statistic is tested using the $\chi^2$ distribution. Replicate and Order effects show evidence, for all three Panels, of a lack of stability. They are important for some variables but not for others. Replicate effects can be due to differences in the Samples - always more of a problem where Samples have to be prepared and then served hot. The Order effects imply a certain lack of familiarity with profiling tomato soup. Clearly more attention needs to be paid by Assessors to the routine between Samples.

The applied treatment effects i.e. Cornflour, Flavour and their interaction, show only sporadic effects. In particular Cornflour appears to have a massive textural effect on the Samples. As far as can be determined this effect is drowning out other more subtle effects.

With the benefit of hindsight, it is apparent that spiked samples must be constructed in a more sophisticated way if the profiling competence of the panels is to be tested.

## Table 1: Results for Panel Y

### (a) Components of Variance

| No. | Variate | Modality | Assessor | Assessor x Sample | Assessor x Session | Assessor x Order | Residual |
|---|---|---|---|---|---|---|---|
| 1 | Intensity | (O) | 43 | 0 | 0 | 1 | 40 |
| 2 | Sour | (O) | 149 | 7 | 0 | 13 | 38 |
| 3 | Tomato-like | (O) | 35 | 0 | 3 | 9 | 26 |
| 4 | Mixed vegetables | (O) | 145 | 0 | 0 | 2 | 42 |
| 5 | Spicy | (O) | 194 | 0 | 13 | 2 | 38 |
| 6 | Thickness | (Te) | 83 | 30 | 26 | 0 | 142 |
| 7 | Intensity | (F) | 83 | 0 | 3 | 4 | 25 |
| 8 | Sweet | (Ta) | 103 | 0 | 0 | 9 | 50 |
| 9 | Acid | (Ta) | 196 | 4 | 12 | 8 | 37 |
| 10 | Salty | (Ta) | 252 | 1 | 2 | 4 | 36 |
| 11 | Tomato-like | (F) | 44 | 0 | 0 | 1 | 28 |
| 12 | Mixed vegetables | (F) | 118 | 5 | 2 | 2 | 25 |
| 13 | Sharp | (F) | 136 | 15 | 13 | 5 | 47 |
| 14 | Spicy | (F) | 195 | 0 | 14 | 9 | 27 |
| 15 | Pungent | (At) | 285 | 0 | 15 | 5 | 70 |

O  – Odour            Te   – Texture
F  – Flavour          Ta   – Taste
At – After Taste

### (b) Wald tests of fixed effects

| Variate | Replicate | Order | Cornflour | Flavour | Cornflour x Flavour |
|---|---|---|---|---|---|
| df | 2 | 2 | 1 | 2 | 2 |
| 1 | 2.3 | 32.2 | 3.4 | 0.2 | 0.0 |
| 2 | 8.4 | 17.5 | 1.0 | 2.4 | 0.2 |
| 3 | 4.2 | 8.5 | 0.1 | 1.7 | 0.5 |
| 4 | 0.3 | 2.0 | 0.2 | 0.9 | 2.9 |
| 5 | 1.5 | 1.5 | 0.0 | 9.3 | 5.2 |
| 6 | 1.4 | 1.1 | 142.9 | 2.7 | 0.8 |
| 7 | 0.4 | 0.5 | 6.5 | 6.9 | 4.5 |
| 8 | 0.5 | 6.4 | 6.8 | 3.6 | 0.8 |
| 9 | 0.1 | 0.3 | 0.8 | 5.4 | 2.2 |
| 10 | 2.4 | 0.8 | 1.7 | 1.2 | 0.9 |
| 11 | 2.6 | 1.8 | 0.6 | 1.6 | 2.6 |
| 12 | 5.0 | 14.0 | 1.3 | 0.0 | 1.1 |
| 13 | 1.1 | 5.8 | 0.1 | 8.6 | 1.0 |
| 14 | 0.8 | 2.6 | 0.3 | 13.9 | 2.2 |
| 15 | 0.1 | 16.9 | 0.1 | 11.1 | 0.1 |

## Table 2:    Results for Panel Z

### (a)  Components of Variance

| No. | Variate | Modality | Assessor | Assessor x Sample | Assessor x Session | Assessor x Order | Residual |
|---|---|---|---|---|---|---|---|
| 1 | Intensity | (O) | 104 | 0 | 2 | 0 | 28 |
| 2 | Tomato | (O) | 142 | 11 | 2 | 0 | 29 |
| 3 | Acid | (O) | 76 | 2 | 6 | 0 | 28 |
| 4 | Herbs | (O) | 19 | 0 | 0 | 3 | 36 |
| 5 | Tomato | (F) | 129 | 1 | 0 | 0 | 30 |
| 6 | Acid | (F) | 131 | 0 | 3 | 5 | 44 |
| 7 | Herbs | (F) | 41 | 2 | 0 | 0 | 22 |
| 8 | Peper | (F) | 112 | 10 | 4 | 1 | 24 |
| 9 | Salt | (F) | 77 | 2 | 3 | 0 | 15 |
| 10 | Sweet | (F) | 160 | 0 | 1 | 0 | 25 |
| 11 | Intensity | (F) | 64 | 0 | 5 | 0 | 24 |
| 12 | Thickness | (Te) | 43 | 17 | 1 | 0 | 29 |
| 13 | Roughness | (Te/Mo) | 42 | 0 | 2 | 0 | 28 |
| 14 | Intensity | (At) | 85 | 0 | 4 | 6 | 28 |
| 15 | Tomato | (At) | 94 | 0 | 0 | 0 | 49 |
| 16 | Acid | (At) | 60 | 4 | 9 | 0 | 45 |
| 17 | Pepper | (At) | 100 | 5 | 13 | 5 | 40 |

O  – Odour

F  – Flavour

Te/Mo – Texture/Mouthfeel

Te  – Texture

At  – After Taste

### (b)  Wald tests of fixed effects

| Variate | Replicate | Order | Cornflour | Flavour | Cornflour x Flavour |
|---|---|---|---|---|---|
| df | 2 | 2 | 1 | 2 | 2 |
| 1 | 2.8 | 1.9 | 10.3 | 3.2 | 0.5 |
| 2 | 0.6 | 3.4 | 3.4 | 0.0 | 2.0 |
| 3 | 1.4 | 0.4 | 4.6 | 0.1 | 1.2 |
| 4 | 0.3 | 0.3 | 2.1 | 1.6 | 0.4 |
| 5 | 1.8 | 1.8 | 11.0 | 2.0 | 0.5 |
| 6 | 0.5 | 0.7 | 0.3 | 9.7 | 0.6 |
| 7 | 1.8 | 2.1 | 0.0 | 2.4 | 0.9 |
| 8 | 9.1 | 0.1 | 0.1 | 2.5 | 0.2 |
| 9 | 2.8 | 0.3 | 0.0 | 3.6 | 0.4 |
| 10 | 0.2 | 4.6 | 0.1 | 7.6 | 3.0 |
| 11 | 1.4 | 0.4 | 1.3 | 3.0 | 0.8 |
| 12 | 6.5 | 1.3 | 179.7 | 0.8 | 0.7 |
| 13 | 1.3 | 0.5 | 2.5 | 0.3 | 0.2 |
| 14 | 1.7 | 0.5 | 1.2 | 2.1 | 1.3 |
| 15 | 1.3 | 0.6 | 3.8 | 4.9 | 2.7 |
| 16 | 1.6 | 3.4 | 2.1 | 2.4 | 0.7 |
| 17 | 0.9 | 6.1 | 0.0 | 3.5 | 0.6 |

## Table 3:   Results for Panel R

### (a)  Components of Variance

| No. | Variate | Modality | Assessor | Assessor x Sample | Assessor x Session | Assessor x Order | Residual |
|-----|---------|----------|----------|-------------------|--------------------|------------------|----------|
|     |         |          |          |                   |                    |                  |          |
| 1 | Intensity | (O) | 34 | 4 | 3 | 4 | 38 |
| 2 | Tomato | (O) | 26 | 3 | 0 | 7 | 44 |
| 3 | Vegetable | (O) | 68 | 0 | 0 | 10 | 28 |
| 4 | Herbs | (O) | 44 | 0 | 0 | 3 | 35 |
| 5 | Sweet | (O) | 74 | 2 | 1 | 0 | 17 |
| 6 | Intensity | (F) | 63 | 2 | 5 | 0 | 32 |
| 7 | Tomato | (F) | 62 | 3 | 0 | 0 | 38 |
| 8 | Sweet | (F) | 24 | 7 | 4 | 0 | 17 |
| 9 | Salty | (Ta) | 42 | 0 | 0 | 0 | 14 |
| 10 | Herbs | (F) | 33 | 1 | 0 | 0 | 33 |
| 11 | Vegetable | (F) | 52 | 0 | 0 | 0 | 25 |
| 12 | Sour | (F) | 17 | 0 | 9 | 7 | 23 |
| 13 | Spicy | (At) | 48 | 0 | 4 | 9 | 70 |
| 14 | Mouthfeel | (Te) | 29 | 0 | 0 | 0 | 134 |
| 15 | Texture | (Te) | 16 | 0 | 0 | 0 | 130 |

O   – Odour      Te   – Texture  
F    – Flavour     Ta   – Taste  
At  – After Taste

### (b)  Wald tests of fixed effects

| Variate | Replicate | Order | Cornflour | Flavour | Cornflour x Flavour |
|---------|-----------|-------|-----------|---------|---------------------|
| df | 2 | 2 | 1 | 2 | 2 |
|    |   |   |   |   |   |
| 1 | 7.9 | 13.0 | 8.0 | 1.8 | 3.6 |
| 2 | 5.1 | 17.5 | 7.9 | 2.7 | 3.2 |
| 3 | 0.8 | 5.8 | 5.1 | 0.7 | 0.2 |
| 4 | 4.6 | 0.8 | 0.5 | 0.9 | 1.4 |
| 5 | 8.6 | 8.1 | 4.8 | 0.1 | 2.4 |
| 6 | 4.2 | 4.2 | 5.9 | 4.3 | 3.6 |
| 7 | 0.3 | 3.6 | 3.7 | 7.5 | 5.5 |
| 8 | 2.5 | 0.3 | 1.1 | 0.8 | 0.6 |
| 9 | 2.9 | 0.4 | 1.2 | 2.9 | 2.2 |
| 10 | 0.2 | 1.0 | 0.7 | 6.4 | 0.4 |
| 11 | 0.8 | 2.2 | 2.3 | 0.6 | 1.1 |
| 12 | 0.8 | 0.6 | 0.4 | 2.1 | 0.2 |
| 13 | 5.3 | 18.4 | 0.6 | 0.6 | 0.8 |
| 14 | 30.0 | 11.5 | 445.7 | 2.3 | 13.6 |
| 15 | 30.5 | 7.0 | 432.3 | 3.9 | 6.8 |

# APPENDIX 1: ATTRIBUTES ABBREVIATIONS AND DEFINITIONS

## Common Attributes

Attributes marked * are common to all panels.

Total strength of odour
Total strength of flavour
Thickness (mouthfeel)
Sweet
Salty

## Scales (convert to 0 – 100)

| Panel | Scale | | Panel | Scale |
|-------|-------|---|-------|-------|
| N | 1 - 9 | | U | 1 – 9 |
| P | 0 – 100 | | V | 0-15 |
| Q | 0 – 100 | | W | 1 – 9 |
| R | 0 – 100 | | Y | 0 – 10 |
| T | 1 -10 | | Z | 0 – 10 |

To convert the scales to all range from 0 to 100, the following procedure should be used.

Let the current scale range from a to b, and the target scale range from A to B.
In this case A = 0 and B = 100.

Let X denote the original value and Y denote the target value.

$$Y = \frac{[(B - A)*X + A*b - B*a]}{b - a}$$

For example, for Panel N, the formula would reduce to the following.

$$Y = \frac{[(100 - 0)*X + (0*9) - (100*1)]}{9 - 1}$$

$$Y = \frac{100X + 0 - 100}{8}$$

$$Y = \frac{100(X - 1)}{8}$$

**Panel N**

| Number | Category | Attribute | Abbreviation | Definition |
|--------|----------|-----------|--------------|------------|
| 1* | Odour | Intensity | O-TOTAL | Smell, no matter what type |
| 2 | Odour | Fresh | O-FRESH | Green smell, fresh, natural |
| 3* | Flavour | Total intensity | F-TOTAL | Taste, no matter what type |
| 4 | Flavour | Tomato | F-TOMATO | |
| 5 | Flavour | Fresh | F-FRESH | Green taste, fresh, natural |
| 6 | Flavour | Vegetables | F-VEG | For all vegetables (inc. onions) |
| 7 | Taste | Acid | ACID | |
| 8 | Flavour | Spicy | F-SPICY | Pepper, herbs, spices, condiments |
| 9* | Taste | Salty | SALTY | |
| 10* | Taste | Sweet | SWEET | |
| 11 | Flavour | Preserve | F-PRESERVE | Describes a special taste coming from the transformation of the product during the conservation (metal, tomato paste, cooked, etc.) |
| 12 | Flavour | Fat | F-FAT | Fat (oil, meat's fat, boiled beef) |
| 13 | Taste | Bitter | BITTER | |
| 14* | Mouthfeel | Thickness | THICK | |
| 15 | Mouthfeel | Smoothness | SMOOTH | Sensation covering the tongue |
| 16 | Mouthfeel | Stickiness | STICKY | |
| 17 | Mouthfeel | Granular | GRANULAR | Mix the soup in the mouth; describes the presence of bumps in the mouth (pulp, pieces of vegetables) |
| 18 | Mouthfeel | Fat | FAT | Oily sensation in the mouth |
| 19 | Mouthfeel | Rough | ROUGH | Which itches in the mouth |
| n/a | Appearance | Deposit on the cup | n/a | |
| n/a | Appearance | Particles present | n/a | |
| n/a | Appearance | Bound | n/a | |
| n/a | Appearance | Gelatinous | n/a | |
| 20 | Aftertaste | Prickly | AT-PRICKLE | Whatever the nature of the prickle (coming from tomatoes or spices) |

**Note:** Appearance attributes are **not** permitted, so discard from any analysis.

**Panel P**

| Number | Category | Attribute | Abbreviation | Definition |
|---|---|---|---|---|
| 1* | Odour | Overall | O-TOTAL | Overall strength of odour |
| 2 | Odour | Tomato | O-TOMATO | Strength of tomato odour, reminiscent of canned tomato soup. |
| 3 | Odour | Herbs | O-HERBS | Strength of herbs odour, reminiscent of dried basil and oregano |
| 4 | Odour | Vegetable Mix | O-VEG | Strength of vegetable mix (e.g. carrot, celery, etc.) |
| n/a | Odour | Other | -- | Any other odour |
| 5* | Flavour | Overall | F-TOTAL | Overall strength of flavour |
| 6 | Flavour | Tomato | F-TOMATO | Strength of tomato flavour, reminiscent of canned tomato soup |
| 7* | Taste | Sweet | SWEET | Intensity of sweet primary taste, associated to sucrose and any other sweetener |
| 8* | Taste | Salty | SALY | Intensity of salty primary taste, associated to salt |
| 9 | Taste | Acidic | ACID | Intensity of acidic primary taste, associated to citric acid |
| 10 | Flavour | Herbs/Spices | F-HERBS/SPICY | Strength of herbs and spices flavour, reminiscent of dried basil, oregano and onion |
| 11 | Flavour | Vegetable Mix | F-VEG | Strength of vegetable mix and celery salt and MSG in particular |
| 12 | Flavour | Stock Cube/MSG | F-MSG | Strength of cornflour flavour, reminiscent of uncooked flour flavour |
| 13 | Flavour | Cornflour | F-FLOUR | Strength of stock cube flavour, reminiscent of MSG |
| n/a | Flavour | Other | -- | Any other flavour |
| 14* | Mouthfeel | Thickness | THICK | Viscosity of soup in mouth |
| 15 | Mouthfeel | Powdery | POWDERY | Mouthfeel associated with the presence of thin particles |
| 16 | Mouthfeel | Teeth Coating | TEETHC | Sticking to teeth after swallowing |
| n/a | Mouthfeel | Other | | Any other texture/mouthfeel |
| 17 | After Taste | Burning/Harsh | AT-HARSH | Sensation reminiscent of spices / herbs after swallowing, harsh, burning mouth/throat |

'Other' terms should be omitted from the analysis as it is not defined.

**Panel Q**

| Number | Category | Attribute | Abbreviation | Definition |
|--------|----------|-----------|--------------|------------|
| 1* | Odour | Total intensity | O-TOTAL | Smell the sample and evaluate the intensity of the odour of the sample before tasting it. |
| 2* | Mouthfeel | Thickness | THICK | Evaluate the thickness of the sample by mouthfeel. |
| 3 | Mouthfeel | Sliminess (gel-like) | SLIMEY | Evaluate the sliminess (gel-likeness) of the sample by mouthfeel. |
| 4 | Mouthfeel | Smoothness | SMOOTH | Evaluate the smoothness by pressing the soup with your tongue against the roof of your mouth. |
| 5* | Flavour | Total intensity | F-TOTAL | Taste the sample and evaluate the total flavour of the sample. |
| 6* | Taste | Sweetness | SWEET | Sweetness is a primary taste which is caused by different kinds of sugars (e.g. saccharine). |
| 7* | Taste | Saltiness | SALTY | Saltiness is a primary taste which is caused by different kinds of salts (most clearly by NaCl, or table salt). |
| 8 | Taste | Sourness | SOUR | Sourness is a primary taste which is caused by different kinds of acids (e.g. citric and lactic acid). |
| 9 | Taste | Umani | UMAMI | Umami is a primary taste which is caused by for example sodium glutamate. |
| 10 | Flavour | Spiciness (Pepper) | F-PEPPER | Peppery spiciness is caused by different kinds of chilli's and peppers and it is recognised as a burning sensation. |
| 11 | Flavour | Herb-like | F-HERBS | This taste is caused by different kinds of herbs, e.g. basil, oregano, thyme, parsley, etc. |
| 12 | Flavour | Tomato-like | F-TOMATO | This taste is caused by the flavour of tomatoes; like fresh tomato is most intense. |
| 13 | Aftertaste | Aftertaste | AFTERT | After-taste is the flavour still left in the mouth after the sample has been spit out. Evaluate the intensity of the after-taste after spitting the sample out. |

**Panel R**

| Number | Category | Attribute | Abbreviation | Definition |
|---|---|---|---|---|
| 1* | Odour | Total | O-TOTAL | Total odour strength |
| 2 | Odour | Tomato | O-TOMATO | Odour of: tomatoes, tomato purée, tomato ketchup and canned tomatoes |
| 3 | Odour | Vegetable soup | O-VEG | Odour of clear vegetable soup |
| 4 | Odour | Herbs | O-HERBS | Odour of basilica and oregano |
| 5 | Odour | Sweet | O-SWEET | Sweet odour |
| 6* | Flavour | Total | F-TOTAL | Total flavour strength |
| 7 | Flavour | Tomato | F-TOMATO | Flavour of: tomatoes, tomato purée, tomato ketchup and canned tomatoes |
| 8* | Taste | Sweet | SWEET | Sweet flavour |
| 9* | Taste | Salty | SALTY | Salty taste |
| 10 | Flavour | Herbs | F-HERBS | Flavour of basilica and oregano |
| 11 | Flavour | Vegetable soup | F-VEG | Flavour of clear vegetable soup |
| 12 | Taste | Sour | SOUR | A mixture of sweet/sour/tomato (marinade) |
| 13 | Aftertaste | Spicy | AT-SPICY | A strong aftertaste similar to pepper |
| 14* | Mouthfeel | Thickness (mouth) | THICK | How thick or thin the soup feels in the mouth |
| 15 | Mouthfeel | Thickness (spoon) | SP-THICK | How thick or thin the soup feels with the spoon |

Note: As Thickness was measured from thick to thin, the data has been converted to go from thin to thick, thus making the attribute in line with other data sets. This saves confusion for analysis.

## Panel T

| Number | Category | Attribute | Abbreviation | Definition |
|---|---|---|---|---|
| 1* | Odour | Overall strength | O-TOTAL | The overall strength of the aroma. |
| 2 | Odour | Tomato | O-TOMATO | The level of tomato aroma . |
| 3 | Odour | Vegetables | O-VEG | The level of vegetable aroma – e.g. onion, carrot, celery, potato etc. Please describe. |
| 4 | Odour | Spicy | O-SPICY | The level of spicy aroma – e.g. pepper, Worcester sauce etc. Please describe. |
| 5 | Odour | Hydrolysate | O-HYDRO | The level of hydrolysate aroma, like stock cube or powdered soup. |
| n/a | Odour | Other | Not used | The level of any other aroma. Please describe. |
| 6* | Flavour | Overall strength | F-TOTAL | The overall strength of the flavour. |
| 7 | Flavour | Tomato | F-TOMATO | The level of tomato flavour. |
| 8 | Flavour | Vegetables | F-VEG | The level of vegetable flavour – e.g. onion, carrot, celery, potato etc. Please describe. |
| 9* | Taste | Sweet | SWEET | The level of sweet flavour. |
| 10* | Taste | Salt | SALTY | The level of salt flavour. |
| 11 | Flavour | Spicy | F-SPICY | The level of spicy flavour –e.g. pepper, Worcester sauce etc. Please describe. |
| 12 | Flavour | Herbs | F-HERBS | The level of herb flavour. Please describe. |
| 13 | Taste | Acidic | ACID | The level of acidic flavour. |
| n/a | Flavour | Others | Not used | The level of any other flavour. |
| 14* | Mouthfeel | Thickness | THICK | The thickness or viscosity of the sample. |
| 15 | Mouthfeel | Gelatinous | GELATIN | The degree to which the sample feels like unset jelly in the mouth. |
| 16 | Mouthfeel | Powdery | POWDERY | The degree to which the sample feels powdery in the mouth. This is assessed by rubbing the tongue across the palate. |
| 17 | Mouthfeel | Smoothness | SMOOTH | The degree to which the sample is perceived to be free of any particles or roughness |
| 18 | Mouthfeel | Mouthcoating | MOUTHC | The amount of coating of fat, powder or gum left around the inside of the mouth after the sample has been expelled. |
| 19 | Mouthfeel | Astringent | ASTRINGENT | The degree to which a puckered, drying sensation is perceived on the sides of the mouth after the sample has been expelled. |
| 20 | Aftertaste | Tomato | AT-TOMATO | The level of tomato aftertaste. |
| 21 | Aftertaste | Sweet | AT-SWEET | The level of sweet aftertaste. |
| 22 | Aftertaste | Salt | AT-SALTY | The level of salt aftertaste. |
| 23 | Aftertaste | Pepper | AT-PEPPER | The level of pepper aftertaste. |
| 24 | Aftertaste | Acidic | AT-ACID | The level of acidic aftertaste. |
| n/a | Aftertaste | Others | Not used | The level of any other aftertaste. |

*'Other' terms should be omitted from the analysis as it is not defined.*

**Panel U**

| Number | Category | Attribute | Abbreviation | Definition |
|---|---|---|---|---|
| 1* | Odour | Total Strength | O-TOTAL | Overall intensity of odour. |
| 2 | Odour | Sweetish | O-SWEET | The strength of sweet odour. |
| 3 | Odour | Tomato Puree | O-TOMATO | The strength of canned tomato purée odour. |
| 4 | Odour | Spicy | O-SPICY | The strength of spicy odour. |
| 5* | Taste | Sweet | SWEET | Intensity of sweet taste associated to sucrose. |
| 6* | Taste | Salty | SALTY | Intensity of salty taste associated to sodium chloride. |
| 7 | Taste | Acid | ACID | Intensity of sour/acid associated to citric acid. |
| 8* | Flavour | Total Strength | F-TOTAL | Overall intensity of flavour. |
| 9 | Flavour | Tomato Puree | F-TOMATO | The strength of canned tomato purée flavour. |
| 10 | Flavour | Spicy | F-SPICY | The strength of spicy flavour |
| 11 | Mouthfeel | Pungent | PUNGENT | Feeling factor associated with paprika pepper. |
| 12 | Mouthfeel | Viscosity | VISCOSE | The force to draw between leaps from glass. |
| 13* | Mouthfeel | Thickness | THICK | The mouthfeel of product thickness (tactile). |

## Panel V

| Number | Category | Attribute | Abbreviation | Definitions |
|--------|----------|-----------|--------------|-------------|
| 1* | Odour | Intensity | O-TOTAL | |
| 2 | Odour | Sour | O-SOUR | |
| 3 | Odour | Sweet | O-SWEET | |
| 4 | Odour | Tomato | O-TOMATO | |
| 5 | Odour | Green soup herbs | O-HERBS-GREEN | |
| 6 | Odour | Soup | O-SOUP | |
| 7 | Odour | Other herbs | O-HERBS | |
| 8* | Flavour | Intensity | F-TOTAL | |
| 9* | Taste | Sweet | SWEET | |
| 10 | Taste | Sour | SOUR | |
| 11* | Taste | Salt | SALTY | |
| 12 | Flavour | Pepper | F-PEPPER | |
| 13 | Flavour | Creamy/butter | F-CREAMY | |
| 14 | Flavour | Tomato | F-TOMATO | |
| 15 | Flavour | Soup | F-SOUP | |
| 16 | Flavour | Garden herbs | F-GHERBS | |
| 17 | Flavour | Artificial | F-ARTICIAL | |
| 18 | Flavour | Thickening agent | F-THAGENT | |
| 19* | Mouthfeel | Thin – Thick | THICK | |
| 20 | Mouthfeel | Soft | SOFT | |
| 21 | Mouthfeel | Lumpy | LUMPY | |
| 22 | Mouthfeel | Mealy | MEALY | |
| 23 | Mouthfeel | Sticky | STICKY | |
| 24 | Aftertaste | Intensity | AT-TOTAL | |
| 25 | Aftertaste | Sour | AT-SOUR | |
| 26 | Aftertaste | Salt | AT-SALTY | |
| 27 | Aftertaste | Sweet | AT-SWEET | |
| 28 | Aftertaste | Burning | AT-BURN | |
| 29 | Aftertaste | Thirst stimulating | AT-THIRST | |
| 30 | Aftertaste | Astringent | AT-ASTRINGENT | |

*No definitions provided*

**Panel W**

| Number | Category | Attribute | Abbreviation | Definition |
|---|---|---|---|---|
| 1* | Odour | Total strength | O-TOTAL | Total strength of all odours. |
| 2 | Odour | Bouillon | O- BOUILLON | Flavour of Bouillon. |
| 3 | Odour | Tomato | O-TOMATO | Flavour of tomato. |
| 4* | Flavour | Total strength | F-TOTAL | Total strength of all flavours. |
| 5 | Flavour | Bouillon | F-BOUILLON | Flavour of Bouillon powder. |
| 6 | Flavour | Tomato | F-TOMATO | Flavour of tomato. |
| 7* | Taste | Sweetness | SWEET | Sweet taste. |
| 8 | Taste | Sourness | SOUR | Sour taste. |
| 9* | Taste | Salty | SALTY | Salty taste. |
| 10 | Taste | Bitter | BITTER | Bitter taste. |
| 11 | Flavour | Pepper | F-PEPPER | Flavour of pepper. |
| 12 | Flavour | Spices | F-SPICY | Flavour of spices (except pepper). |
| 13 | Flavour | Flour | F-FLOUR | Flavour of flour, grain. |
| 14 | Flavour | Rancid | F-RANCID | Flavour of oxidised lipids. |
| 15 | Mouthfeel | Smoothness | SMOOTH | Smooth, even texture. |
| 16* | Mouthfeel | Thickness | THICK | Viscosity, mouthfeel. |
| 17 | Aftertaste+ | Aftertaste (30s) | AFTERT | Flavour measured 30 seconds after the sample has been spitted from the mouth. |

+ after 30 seconds

## Panel Y

| Number | Category | Attribute | Abbreviation | Definition |
|---|---|---|---|---|
| 1* | Odour | Total intensity | O-TOTAL | Overall perception of odour intensity (strength) of the sample. |
| 2 | Odour | Sour | O-SOUR | Typical for tomato puree concentrate (freshly open can). |
| 3 | Odour | Tomato-like | O-TOMATO | Characteristic for cooked tomatoes . |
| 4 | Odour | Mixed vegetable-like | O-VEG | Characteristic for freshly cooked and mixed vegetables. Homogenised. |
| 5 | Odour | Spicy | O-SPICY | Characteristic for spices mixture (with the pronounced maggi note). |
| 6* | Mouthfeel | Thickness | THICK | Perception of degree of thickness in the mouth. |
| 7* | Flavour | Total intensity | F-TOTAL | Overall perception of flavour intensity of the sample. |
| 8* | Taste | Sweet | SWEET | Basic taste. |
| 9 | Taste | Acid | ACID | Basic taste. |
| 10* | Taste | Salty | SALTY | Basic taste. |
| 11 | Flavour | Tomato-like | F-TOMATO | Characteristic for cooked tomatoes . |
| 12 | Flavour | Mixed vegetable-like | F-VEG | Characteristic for freshly cooked and homogenised mixed vegetables. |
| 13 | Flavour | Sharp, irritating | F-SHARP | Irritating sensation in the mouth |
| 14 | Flavour | Spicy | F-SPICY | Characteristic for spices mixture (with the pronounced maggi note). |
| 15 | Aftertaste | Pungent | AT-PUNGENT | Feeling of pungency on the edges of tongue, appearing as after-taste, long-lasting. |

## Panel Z

| Number | Category | Attribute | Abbreviation | Definition |
|--------|----------|-----------|--------------|------------|
| 1* | Odour | Total intensity | O-TOTAL | The intensity of all odours present in the sample. |
| 2 | Odour | Tomato | O-TOMATO | The intensity of tomato odour. |
| 3 | Odour | Acid | O-ACID | The intensity of acid odour. |
| 4 | Odour | Herb | O-HERBS | The intensity of herb odours. |
| 5* | Flavour | Total intensity | F-TOTAL | The intensity of all flavours present in the sample. |
| 6 | Flavour | Tomato | F-TOMATO | The intensity of tomato flavour. |
| 7 | Flavour | Acid | ACID | The intensity of acid taste. |
| 8 | Flavour | Herb | F-HERBS | The intensity of herb flavours. |
| 9 | Flavour | Pepper | F-PEPPER | The intensity of pepper flavour. |
| 10* | Flavour | Salt | SALTY | The intensity of salt taste. |
| 11* | Flavour | Sweet | SWEET | The intensity of sweet taste. |
| 12 | Mouthfeel | Thick | THICK | The thick sensation felt between the tongue and palate immediately after taking the soup into the mouth. |
| 13 | Mouthfeel | Rough | ROUGH | The rough sensation felt in the mouth when eating the soup. |
| 14 | Aftertaste | Total intensity | AT-TOTAL | The intensity of all remaining flavours left 10 to 15 seconds after the sample is expectorated . |
| 15 | Aftertaste | Tomato | AT-TOMATO | The intensity of remaining tomato flavour. |
| 16 | Aftertaste | Acid | AT-ACID | The intensity of remaining acid taste. |
| 17 | Aftertaste | Pepper | AT-PEPPER | The intensity of remaining pepper flavour. |

## APPENDIX 2: MEAN PANEL DATA FOR THE 5 COMMON ATTRIBUTES

| Panel | Sample | ATTRIBUTES | | | | |
|-------|--------|---------|---------|-------|-------|-------|
| | | O-TOTAL | F-TOTAL | SWEET | SALTY | THICK |
| N | 219 | 5.7 | 5.3 | 5.6 | 4.0 | 6.2 |
| | 250 | 5.7 | 5.5 | 5.1 | 4.2 | 3.8 |
| | 315 | 5.8 | 5.3 | 5.6 | 4.3 | 6.2 |
| | 591 | 5.9 | 5.3 | 5.6 | 4.2 | 6.0 |
| | 618 | 5.8 | 5.4 | 5.2 | 4.2 | 4.0 |
| | 904 | 5.7 | 5.3 | 5.2 | 4.2 | 3.5 |
| P | 219 | 55.0 | 59.9 | 47.1 | 62.7 | 73.2 |
| | 250 | 58.6 | 56.9 | 45.9 | 62.6 | 53.7 |
| | 315 | 56.9 | 57.9 | 49.2 | 62.4 | 79.6 |
| | 591 | 57.5 | 59.2 | 52.1 | 60.9 | 73.0 |
| | 618 | 62.9 | 61.9 | 51.1 | 62.5 | 56.0 |
| | 904 | 61.3 | 64.8 | 53.1 | 64.3 | 54.9 |
| Q | 219 | 60.8 | 64.0 | 53.1 | 51.7 | 45.6 |
| | 250 | 57.5 | 61.4 | 50.1 | 53.0 | 37.8 |
| | 315 | 56.9 | 61.0 | 49.3 | 52.6 | 47.0 |
| | 591 | 55.3 | 60.7 | 53.8 | 54.7 | 44.7 |
| | 618 | 61.0 | 62.3 | 55.8 | 49.9 | 33.4 |
| | 904 | 59.4 | 62.4 | 52.5 | 52.7 | 37.8 |
| R | 219 | 41.2 | 45.6 | 32.7 | 19.9 | 69.9 |
| | 250 | 45.0 | 47.4 | 30.6 | 20.9 | 32.7 |
| | 315 | 44.1 | 46.8 | 32.5 | 20.8 | 66.0 |
| | 591 | 42.9 | 46.2 | 32.5 | 20.3 | 79.3 |
| | 618 | 46.7 | 47.6 | 31.5 | 20.0 | 31.7 |
| | 904 | 46.6 | 51.1 | 32.5 | 22.2 | 28.8 |
| T | 219 | 5.1 | 6.0 | 5.2 | 4.4 | 6.7 |
| | 250 | 4.9 | 5.7 | 5.2 | 4.4 | 3.5 |
| | 315 | 5.0 | 5.8 | 5.2 | 4.5 | 6.6 |
| | 591 | 5.2 | 6.0 | 5.5 | 4.1 | 6.4 |
| | 618 | 5.0 | 5.4 | 4.6 | 4.1 | 3.4 |
| | 904 | 4.9 | 5.6 | 5.3 | 4.2 | 3.7 |

| Panel | Sample | ATTRIBUTES | | | | |
|---|---|---|---|---|---|---|
| | | O-TOTAL | F-TOTAL | SWEET | SALTY | THICK |
| U | 219 | 6.7 | 6.7 | 6.0 | 4.1 | 7.3 |
| | 250 | 6.2 | 6.1 | 5.0 | 4.5 | 4.4 |
| | 315 | 6.6 | 6.6 | 5.8 | 4.3 | 6.9 |
| | 591 | 6.7 | 6.7 | 6.0 | 4.1 | 7.1 |
| | 618 | 5.8 | 5.9 | 5.1 | 4.6 | 3.7 |
| | 904 | 6.0 | 6.0 | 5.2 | 4.6 | 4.0 |
| V | 219 | 9.2 | 9.9 | 9.1 | 5.8 | 11.5 |
| | 250 | 10.0 | 10.5 | 8.4 | 6.1 | 5.7 |
| | 315 | 9.5 | 9.3 | 8.8 | 5.7 | 11.8 |
| | 591 | 9.3 | 9.3 | 9.3 | 5.0 | 11.6 |
| | 618 | 8.8 | 10.1 | 9.2 | 5.7 | 5.5 |
| | 904 | 9.4 | 9.8 | 8.7 | 6.0 | 5.3 |
| W | 219 | 5.2 | 5.5 | 4.0 | 3.0 | 6.6 |
| | 250 | 5.1 | 5.7 | 4.1 | 3.2 | 4.2 |
| | 315 | 5.3 | 5.8 | 4.1 | 3.0 | 5.4 |
| | 591 | 5.0 | 5.5 | 4.2 | 3.0 | 7.2 |
| | 618 | 5.2 | 5.8 | 4.3 | 3.2 | 4.2 |
| | 904 | 5.4 | 5.6 | 4.2 | 3.0 | 5.0 |
| Y | 219 | 4.6 | 5.4 | 4.0 | 2.5 | 6.5 |
| | 250 | 4.4 | 5.3 | 3.3 | 2.6 | 3.2 |
| | 315 | 4.7 | 5.4 | 3.5 | 2.6 | 6.6 |
| | 591 | 4.6 | 5.3 | 3.9 | 2.6 | 7.0 |
| | 618 | 4.4 | 5.1 | 3.5 | 2.4 | 3.7 |
| | 904 | 4.4 | 4.8 | 3.4 | 2.3 | 3.9 |
| Z | 219 | 6.6 | 6.7 | 7.3 | 3.5 | 5.8 |
| | 250 | 6.5 | 6.3 | 7.1 | 3.1 | 3.7 |
| | 315 | 6.8 | 6.6 | 7.2 | 3.2 | 5.8 |
| | 591 | 6.6 | 6.6 | 7.1 | 3.2 | 5.7 |
| | 618 | 6.4 | 6.4 | 7.2 | 3.4 | 3.9 |
| | 904 | 6.3 | 6.3 | 7.1 | 3.5 | 3.9 |

# APPENDIX 3: ANALYSIS OF VARIANCE P-VALUE RESULTS

The tables below report the results of a mixed model ANOVA (p-values), where assessor was taken as the random effect.

**Panel N**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.782 | 0.531 | 0.958 | 0.397 |
| 2 | O-FRESH | 0.203 | 0.309 | 0.584 | 0.021 |
| 3* | F-TOTAL | 0.836 | 0.490 | 0.726 | 0.648 |
| 4 | F-TOMATO | 0.595 | 0.176 | 0.554 | 0.629 |
| 5 | F-FRESH | 0.058 | 0.007 | 0.112 | 0.267 |
| 6 | F-VEG | 0.263 | 0.061 | 0.356 | 0.293 |
| 7 | ACID | 0.106 | 0.128 | 0.193 | 0.314 |
| 8 | F-SPICY | 0.044 | 0.248 | 0.184 | 0.078 |
| 9* | SALTY | 0.373 | 0.301 | 0.389 | 0.336 |
| 10* | SWEET | 0.031 | 0.000 | 0.952 | 0.863 |
| 11 | F-PROCESS | 0.194 | 0.046 | 0.305 | 0.939 |
| 12 | F-FAT | 0.000 | 0.000 | 0.490 | 0.848 |
| 13 | BITTER | 0.919 | 0.580 | 0.736 | 0.584 |
| 14* | THICK | 0.000 | 0.000 | 0.063 | 0.493 |
| 15 | SMOOTH | 0.000 | 0.000 | 0.370 | 0.602 |
| 16 | STICKY | 0.000 | 0.000 | 0.253 | 0.428 |
| 17 | GRANULAR | 0.011 | 0.001 | 0.197 | 0.740 |
| 18 | FAT | 0.000 | 0.000 | 0.790 | 0.614 |
| 19 | ROUGH | 0.500 | 0.117 | 0.374 | 0.896 |
| 20 | AT-PRICKLE | 0.014 | 0.953 | 0.027 | 0.014 |

**Panel P**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.015 | 0.005 | 0.658 | 0.246 |
| 2 | O-TOMATO | 0.246 | 0.127 | 0.362 | 0.451 |
| 3 | O-HERBS | 0.166 | 0.030 | 0.926 | 0.492 |
| 4 | O-VEG | 0.181 | 0.506 | 0.357 | 0.115 |
| 5* | F-TOTAL | 0.489 | 0.244 | 0.110 | 0.356 |
| 6 | F-TOMATO | 0.067 | 0.014 | 0.044 | 0.591 |
| 7* | SWEET | 0.175 | 0.756 | 0.087 | 0.297 |
| 8* | SALTY | 0.939 | 0.494 | 0.998 | 0.613 |
| 9 | ACID | 0.282 | 0.163 | 0.409 | 0.264 |
| 10 | F-HERBS/SPICY | 0.399 | 0.030 | 0.812 | 0.766 |
| 11 | F-VEG | 0.192 | 0.094 | 0.155 | 0.308 |
| 12 | F-MSG | 0.047 | 0.012 | 0.786 | 0.064 |
| 13 | F-FLOUR | 0.000 | 0.000 | 0.631 | 0.649 |
| 14* | THICK | 0.000 | 0.000 | 0.465 | 0.116 |
| 15 | POWDERY | 0.000 | 0.000 | 0.586 | 0.220 |
| 16 | TEETHC | 0.000 | 0.000 | 0.928 | 0.990 |
| 17 | AT-HARSH | 0.558 | 0.769 | 0.166 | 0.691 |

**Panel Q**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.285 | 0.243 | 0.053 | 0.458 |
| 2* | THICK | 0.002 | 0.000 | 0.416 | 0.486 |
| 3 | SLIMEY | 0.028 | 0.000 | 0.958 | 0.262 |
| 4 | SMOOTH | 0.254 | 0.359 | 0.215 | 0.457 |
| 5* | F-TOTAL | 0.890 | 0.927 | 0.547 | 0.674 |
| 6* | SWEET | 0.170 | 0.619 | 0.026 | 0.523 |
| 7* | SALTY | 0.309 | 0.354 | 0.139 | 0.660 |
| 8 | SOUR | 0.407 | 0.716 | 0.285 | 0.189 |
| 9 | UMAMI | 0.041 | 0.750 | 0.212 | 0.015 |
| 10 | F-PEPPER | 0.402 | 0.547 | 0.309 | 0.457 |
| 11 | F-HERBS | 0.495 | 0.270 | 0.386 | 0.895 |
| 12 | F-TOMATO | 0.341 | 0.469 | 0.813 | 0.067 |
| 13 | AFTERT | 0.111 | 0.962 | 0.221 | 0.077 |

**Panel R**

|        |           |        | Ingredient ANOVA | | |
|--------|-----------|--------|---------|---------|-------|
| Number | Attribute | Sample | Cornflr | Flavour | C*F |
| 1*     | O-TOTAL   | 0.135  | 0.007   | 0.858   | 0.297 |
| 2      | O-TOMATO  | 0.116  | 0.010   | 0.638   | 0.295 |
| 3      | O-VEG     | 0.229  | 0.019   | 0.498   | 0.995 |
| 4      | O-HERBS   | 0.698  | 0.550   | 0.680   | 0.477 |
| 5      | O-SWEET   | 0.204  | 0.026   | 0.777   | 0.294 |
| 6*     | F-TOTAL   | 0.110  | 0.020   | 0.252   | 0.256 |
| 7      | F-TOMATO  | 0.060  | 0.056   | 0.055   | 0.077 |
| 8*     | SWEET     | 0.841  | 0.235   | 0.671   | 0.632 |
| 9*     | SALTY     | 0.127  | 0.259   | 0.202   | 0.374 |
| 10     | F-HERBS   | 0.240  | 0.423   | 0.053   | 0.888 |
| 11     | F-VEG     | 0.516  | 0.151   | 0.906   | 0.569 |
| 12     | SOUR      | 0.987  | 0.839   | 0.920   | 0.796 |
| 13     | AT-SPICY  | 0.954  | 0.442   | 0.998   | 0.847 |
| 14*    | THICK     | 0.000  | 0.000   | 0.201   | 0.005 |
| 15     | SP-THICK  | 0.000  | 0.000   | 0.116   | 0.061 |

Note: As Thickness was measured from thick to thin, the data has been converted to go from thin to thick, thus making the attribute in line with other data sets. This saves confusion for analysis.

**Panel T**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.788 | 0.320 | 0.674 | 0.638 |
| 2 | O-TOMATO | 0.474 | 0.210 | 0.704 | 0.307 |
| 3 | O-VEG | 0.320 | 0.004 | 0.247 | 0.176 |
| 4 | O-SPICY | 0.027 | 0.001 | 0.044 | 0.843 |
| 5 | O-HYDRO | 0.405 | 0.545 | 0.369 | 0.381 |
| 6* | F-TOTAL | 0.124 | 0.013 | 0.841 | 0.274 |
| 7 | F-TOMATO | 0.474 | 0.180 | 0.848 | 0.204 |
| 8 | F-VEG | 0.007 | 0.000 | 0.682 | 0.457 |
| 9* | SWEET | 0.007 | 0.030 | 0.004 | 0.189 |
| 10* | SALTY | 0.241 | 0.169 | 0.076 | 0.339 |
| 11 | F-SPICY | 0.968 | 1.000 | 0.706 | 0.941 |
| 12 | F-HERBS | 0.787 | 0.505 | 0.823 | 0.697 |
| 13 | ACID | 0.419 | 0.042 | 0.476 | 0.860 |
| 14* | THICK | 0.000 | 0.000 | 0.954 | 0.304 |
| 15 | GELATIN | 0.000 | 0.000 | 0.694 | 0.363 |
| 16 | POWDERY | 0.987 | 0.606 | 0.805 | 0.890 |
| 17 | SMOOTH | 0.963 | 0.911 | 0.507 | 0.907 |
| 18 | MOUTHC | 0.000 | 0.000 | 0.949 | 0.933 |
| 19 | ASTRINGENT | 0.374 | 0.112 | 0.554 | 0.544 |
| 20 | AT-TOMATO | 0.035 | 0.002 | 0.958 | 0.515 |
| 21 | AT-SWEET | 0.011 | 0.001 | 0.035 | 0.657 |
| 22 | AT-SALTY | 0.783 | 0.369 | 0.382 | 0.935 |
| 23 | AT-PEPPER | 0.102 | 0.769 | 0.200 | 0.135 |
| 24 | AT-ACID | 0.600 | 0.423 | 0.217 | 0.740 |

**Panel U**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.000 | 0.000 | 0.357 | 0.043 |
| 2 | O-SWEET | 0.000 | 0.000 | 0.300 | 0.019 |
| 3 | O-TOMATO | 0.590 | 0.522 | 0.932 | 0.157 |
| 4 | O-SPICY | 0.000 | 0.000 | 0.765 | 0.079 |
| 5* | SWEET | 0.000 | 0.000 | 0.079 | 0.659 |
| 6* | SALTY | 0.000 | 0.000 | 0.661 | 0.390 |
| 7 | ACID | 0.226 | 0.020 | 0.078 | 0.357 |
| 8* | F-TOTAL | 0.000 | 0.000 | 0.522 | 0.353 |
| 9 | F-TOMATO | 0.002 | 0.000 | 0.173 | 0.422 |
| 10 | F-SPICY | 0.000 | 0.000 | 0.688 | 0.870 |
| 11 | PUNGENT | 0.119 | 0.005 | 0.614 | 0.781 |
| 12 | VISCOSE | 0.000 | 0.000 | 0.957 | 0.073 |
| 13* | THICK | 0.000 | 0.000 | 0.717 | 0.001 |

**Panel V**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.484 | 0.710 | 0.155 | 0.453 |
| 2 | O-SOUR | 0.217 | 0.266 | 0.638 | 0.064 |
| 3 | O-SWEET | 0.526 | 0.907 | 0.460 | 0.494 |
| 4 | O-TOMATO | 0.357 | 0.840 | 0.128 | 0.689 |
| 5 | O-HERBS-GREEN | 0.577 | 0.171 | 0.647 | 0.443 |
| 6 | O-SOUP | 0.443 | 0.259 | 0.786 | 0.146 |
| 7 | O-HERBS | 0.133 | 0.069 | 0.679 | 0.201 |
| 8* | F-TOTAL | 0.137 | 0.036 | 0.416 | 0.373 |
| 9* | SWEET | 0.533 | 0.338 | 0.450 | 0.684 |
| 10 | SOUR | 0.449 | 0.819 | 0.854 | 0.189 |
| 11* | SALTY | 0.394 | 0.089 | 0.384 | 0.173 |
| 12 | F-PEPPER | 0.058 | 0.142 | 0.041 | 0.225 |
| 13 | F-CREAMY | 0.001 | 0.000 | 0.798 | 0.896 |
| 14 | F-TOMATO | 0.848 | 0.227 | 0.481 | 0.908 |
| 15 | F-SOUP | 0.679 | 0.199 | 0.392 | 0.916 |
| 16 | F-GHERBS | 0.171 | 0.364 | 0.033 | 0.499 |
| 17 | F-ARTICIAL | 0.795 | 0.757 | 0.398 | 0.958 |
| 18 | F-THAGENT | 0.000 | 0.000 | 0.673 | 0.456 |
| 19* | THICK | 0.000 | 0.000 | 0.702 | 0.873 |
| 20 | SOFT | 0.000 | 0.000 | 0.982 | 0.620 |
| 21 | LUMPY | 0.000 | 0.000 | 0.964 | 0.817 |
| 22 | MEALY | 0.000 | 0.000 | 0.085 | 0.935 |
| 23 | STICKY | 0.000 | 0.000 | 0.602 | 0.328 |
| 24 | AT-TOTAL | 0.994 | 0.960 | 0.856 | 0.856 |
| 25 | AT-SOUR | 0.657 | 0.966 | 0.297 | 0.456 |
| 26 | AT-SALTY | 0.806 | 0.597 | 0.871 | 0.365 |
| 27 | AT-SWEET | 0.468 | 0.250 | 0.783 | 0.082 |
| 28 | AT-BURN | 0.382 | 0.649 | 0.059 | 0.676 |
| 29 | AT-THIRST | 0.494 | 0.601 | 0.140 | 0.994 |
| 30 | AT-ASTRINGENT | 0.305 | 0.525 | 0.860 | 0.173 |

**Panel W**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.629 | 0.448 | 0.991 | 0.167 |
| 2 | O- BOUILLON | 0.234 | 0.692 | 0.191 | 0.208 |
| 3 | O-TOMATO | 0.581 | 0.264 | 0.448 | 0.489 |
| 4* | F-TOTAL | 0.518 | 0.455 | 0.243 | 0.228 |
| 5 | F-BOUILLON | 0.214 | 0.099 | 0.705 | 0.123 |
| 6 | F-TOMATO | 0.596 | 0.179 | 0.819 | 0.318 |
| 7* | SWEET | 0.508 | 0.451 | 0.715 | 0.352 |
| 8 | SOUR | 0.039 | 0.008 | 0.860 | 0.008 |
| 9* | SALTY | 0.507 | 0.265 | 0.603 | 0.603 |
| 10 | BITTER | 0.499 | 0.626 | 0.267 | 0.802 |
| 11 | F-PEPPER | 0.404 | 0.164 | 0.264 | 0.559 |
| 12 | F-SPICY | 0.866 | 0.925 | 0.607 | 0.833 |
| 13 | F-FLOUR | 0.000 | 0.000 | 0.650 | 0.384 |
| 14 | F-RANCID | 0.834 | 0.386 | 0.832 | 0.667 |
| 15 | SMOOTH | 0.538 | 0.897 | 0.907 | 0.244 |
| 16* | THICK | 0.000 | 0.000 | 0.000 | 0.000 |
| 17 | AFTERT (+) | 0.419 | 0.532 | 0.268 | 0.554 |

+ after 30 seconds

**Panel Y**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.622 | 0.077 | 0.876 | 0.963 |
| 2 | O-SOUR | 0.607 | 0.174 | 0.241 | 0.718 |
| 3 | O-TOMATO | 0.734 | 0.626 | 0.463 | 0.519 |
| 4 | O-VEG | 0.392 | 0.526 | 0.481 | 0.213 |
| 5 | O-SPICY | 0.123 | 0.575 | 0.081 | 0.226 |
| 6* | THICK | 0.000 | 0.000 | 0.178 | 0.628 |
| 7* | F-TOTAL | 0.004 | 0.003 | 0.033 | 0.095 |
| 8* | SWEET | 0.006 | 0.002 | 0.098 | 0.786 |
| 9 | ACID | 0.088 | 0.166 | 0.052 | 0.239 |
| 10* | SALTY | 0.403 | 0.138 | 0.429 | 0.523 |
| 11 | F-TOMATO | 0.477 | 0.417 | 0.436 | 0.323 |
| 12 | F-VEG | 0.611 | 0.113 | 0.823 | 0.414 |
| 13 | F-SHARP | 0.105 | 0.550 | 0.007 | 0.503 |
| 14 | F-SPICY | 0.139 | 0.582 | 0.068 | 0.201 |
| 15 | AT-PUNGENT | 0.412 | 0.737 | 0.063 | 0.828 |

**Panel Z**

| Number | Attribute | Sample | Ingredient ANOVA | | |
|---|---|---|---|---|---|
| | | | Cornflr | Flavour | C*F |
| 1* | O-TOTAL | 0.014 | 0.002 | 0.192 | 0.809 |
| 2 | O-TOMATO | 0.378 | 0.018 | 0.999 | 0.240 |
| 3 | O-ACID | 0.249 | 0.022 | 0.989 | 0.487 |
| 4 | O-HERBS | 0.567 | 0.137 | 0.510 | 0.844 |
| 5* | F-TOTAL | 0.036 | 0.001 | 0.430 | 0.796 |
| 6 | F-TOMATO | 0.050 | 0.477 | 0.003 | 0.743 |
| 7 | ACID | 0.648 | 0.727 | 0.270 | 0.569 |
| 8 | F-HERBS | 0.778 | 0.686 | 0.193 | 0.909 |
| 9 | F-PEPPER | 0.573 | 0.855 | 0.221 | 0.680 |
| 10* | SALTY | 0.064 | 0.812 | 0.013 | 0.237 |
| 11* | SWEET | 0.534 | 0.280 | 0.478 | 0.850 |
| 12 | THICK | 0.000 | 0.000 | 0.517 | 0.548 |
| 13 | ROUGH | 0.608 | 0.099 | 0.852 | 0.850 |
| 14 | AT-TOTAL | 0.474 | 0.236 | 0.415 | 0.550 |
| 15 | AT-TOMATO | 0.041 | 0.058 | 0.132 | 0.301 |
| 16 | AT-ACID | 0.271 | 0.098 | 0.161 | 0.669 |
| 17 | AT-PEPPER | 0.525 | 0.794 | 0.111 | 0.728 |
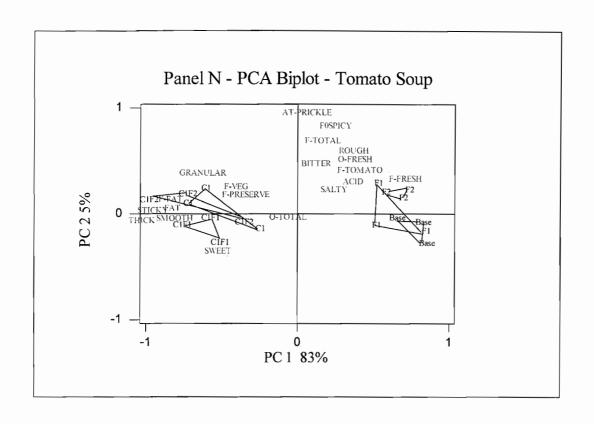
# APPPENDIX 4: SOME RESULTS FROM PCA
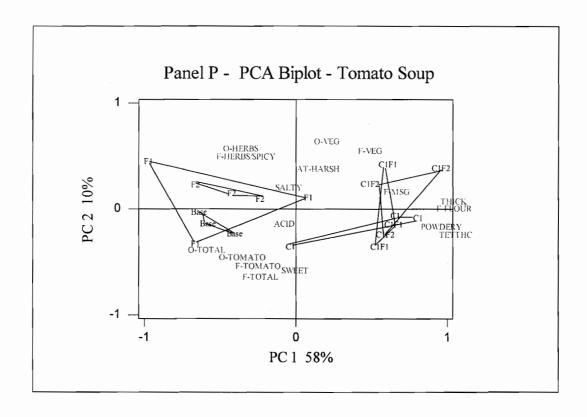
**Percentage Variation Explained by the First 4 PCs**
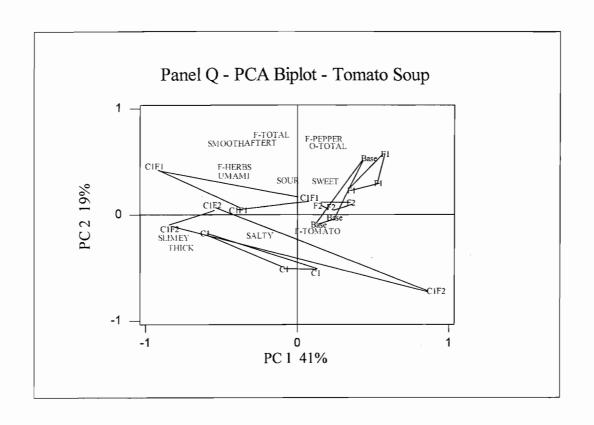
| Panel | PC 1 | PC 2 | PC 3 | PC 4 |
|:-----:|:----:|:----:|:----:|:----:|
| N | 83.0 | 5.1 | 3.3 | 2.6 |
| P | 57.5 | 10.3 | 6.3 | 6.0 |
| Q | 41.3 | 19.1 | 11.4 | 8.1 |
| R | 94.4 | 2.4 | 1.3 | 0.6 |
| T | 86.1 | 3.2 | 2.6 | 1.7 |
| U | 93.9 | 2.9 | 0.7 | 0.6 |
| V | 70.9 | 6.3 | 4.9 | 3.4 |
| W | 59.7 | 22.6 | 6.5 | 3.6 |
| Y | 78.6 | 8.1 | 4.4 | 2.5 |
| Z | 68.4 | 8.8 | 5.2 | 3.9 |

**PCA Biplots**

The following 5 pages show the two dimensional PCA sample and attribute biplots derived from PCA.

## Panel N - PCA Biplot - Tomato Soup



## Panel P - PCA Biplot - Tomato Soup

## Panel Q - PCA Biplot - Tomato Soup



Panel Q - PCA Biplot - Tomato Soup

(Plot axes: PC 1 41% horizontal; PC 2 19% vertical)

Labels visible: F-TOTAL, SMOOTHAFTERT, F-PEPPER, O-TOTAL, Base, F1, C1F1, F-HERBS, UMAMI, SOUR, SWEET, F1, C1F1, F2, F2, F2, Base, Base, C1F2, C1F1, F-TOMATO, C1F2, SLIMEY, THICK, C1, SALTY, C1, C1, C1F2

## Panel R - PCA Biplot - Tomato Soup



Panel R - PCA Biplot - Tomato Soup

(Plot axes: PC 1 94% horizontal; PC 2 2% vertical)

Labels visible: F-HERBS, AT-SPICY, SOUR, F-TOMATO, O-HERBS, F-TOTAL, O-TOTAL, SALTY, O-TOMATO, O-VEG, SWEET, F-VEG, O-SWEET, C1, C1, C1F2, F2, Base, Base, Base, THICK, C1F1, C1F2, F2, F2, SP-THICK, C1, C1F1, F1, F1, C1F2, C1F1, F2

Panel T - PCA Biplot - Tomato Soup



Panel U - PCA Biplot - Tomato Soup

## Panel V - PCA Biplot - Tomato Soup

PC 2 6%

PC 1 71%

F-ARTICIAL  AT-SALTY  F-PEPPER
O-HERBS  AT-BURN  AT-ASTRINGENT
CIF1  CIF1  O-TOTAL  Base
F-THAGENT  CIF2  O-SWEET  F2
MEALY  O-SOUP  F-SOUP  F2
STICKY  AT-THIRST  F1
LUMPY  CIF2
THICK  F-CREAMY  F1  F1
SOUR
SOFT  CIF4  CI  O-HERBS-GREEN  AT-TOTAL  Base
CIF1  AT-SWEET  O-TOMATO
AT-SWEET  AT-SOUR
F-GHERBS  AT-SOUR
O-SOUR  F-TOTAL
SWEET  F-TOMATO

## Panel W - PCA Biplot - Tomato Soup

PC 2 23%

PC 1 60%

AFTERT
F-PEPPER  SALTY
BITTER  F-SPICY
F-RANCID  F-BOUILLON
CIF2  F-BOUILLON  E1
F-FLOUR  F2
O-BOUILLON  Base  TOMATO
SWEET  O-TOMATO
CIF1  F-TOTAL  SOUR
CI  CIF2  F
THICK  Base  O-TOTAL  F1  F2
CIF1  Base  F2
CI  Base
CI  CIF1
SMOOTH

# Panel Y - PCA Biplot - Tomato Soup



# Panel Z - PCA Biplot - Tomato Soup

# APPENDIX 5: GPA SAMPLE PLOTS



Panel N - Consensus GPA Sample Map - Tomato Soup



Panel P - Consensus GPA Sample Map - Tomato Soup

# Panel Q - Consensus GPA Sample Map - Tomato Soup



Dim 2 16%

Dim 1 24%

# Panel R - Consensus GPA Sample Map - Tomato Soup



Dim 2 4%

Dim 1 83%

Panel T - Consensus GPA Sample Map - Tomato Soup



Panel U - Consensus GPA Sample Map - Tomato Soup

# Panel V - Consensus GPA Sample Map - Tomato Soup



# Panel W - Consensus GPA Sample Map - Tomato Soup

## Panel Y - Consensus GPA Sample Map - Tomato Soup

Dim 2 12%

Dim 1 43%

## Panel Z - Consensus GPA Sample Map - Tomato Soup

Dim 2 11%

Dim 1 30%

# APPPENDIX 6: RV COEFFICIENTS FROM GPA RESULTS

## RV Between Assessors in a Panel – All Attributes

The following tables show the RV coefficients calculated to measure the agreement between assessors in each of the 10 sensory panels.

## Panel N – 15 Assessors

|    | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---|
| 1  | 1 | | | | | | | | | | | | | | |
| 2  | 0.68 | 1 | | | | | | | | | | | | | |
| 3  | 0.47 | 0.53 | 1 | | | | | | | | | | | | |
| 4  | 0.45 | 0.47 | 0.55 | 1 | | | | | | | | | | | |
| 5  | 0.41 | 0.46 | 0.46 | 0.49 | 1 | | | | | | | | | | |
| 6  | 0.53 | 0.55 | 0.53 | 0.48 | 0.4 | 1 | | | | | | | | | |
| 7  | 0.7 | 0.57 | 0.55 | 0.56 | 0.49 | 0.58 | 1 | | | | | | | | |
| 8  | 0.68 | 0.66 | 0.47 | 0.51 | 0.38 | 0.42 | 0.63 | 1 | | | | | | | |
| 9  | 0.78 | 0.6 | 0.56 | 0.54 | 0.39 | 0.6 | 0.69 | 0.66 | 1 | | | | | | |
| 10 | 0.49 | 0.63 | 0.38 | 0.39 | 0.46 | 0.39 | 0.51 | 0.47 | 0.46 | 1 | | | | | |
| 11 | 0.38 | 0.44 | 0.33 | 0.34 | 0.32 | 0.37 | 0.4 | 0.45 | 0.38 | 0.34 | 1 | | | | |
| 12 | 0.77 | 0.65 | 0.46 | 0.47 | 0.49 | 0.56 | 0.69 | 0.56 | 0.72 | 0.52 | 0.38 | 1 | | | |
| 13 | 0.59 | 0.59 | 0.51 | 0.56 | 0.44 | 0.51 | 0.62 | 0.61 | 0.61 | 0.53 | 0.41 | 0.67 | 1 | | |
| 14 | 0.84 | 0.7 | 0.52 | 0.49 | 0.42 | 0.56 | 0.66 | 0.66 | 0.75 | 0.57 | 0.46 | 0.73 | 0.58 | 1 | |
| 15 | 0.64 | 0.7 | 0.36 | 0.44 | 0.39 | 0.51 | 0.57 | 0.54 | 0.62 | 0.59 | 0.34 | 0.72 | 0.62 | 0.69 | 1 |

## Panel P – 9 Assessors

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | | |
| 2 | 0.52 | 1 | | | | | | | |
| 3 | 0.29 | 0.55 | 1 | | | | | | |
| 4 | 0.29 | 0.48 | 0.54 | 1 | | | | | |
| 5 | 0.33 | 0.44 | 0.44 | 0.49 | 1 | | | | |
| 6 | 0.5 | 0.65 | 0.51 | 0.42 | 0.49 | 1 | | | |
| 7 | 0.23 | 0.43 | 0.44 | 0.54 | 0.5 | 0.58 | 1 | | |
| 8 | 0.3 | 0.39 | 0.38 | 0.51 | 0.57 | 0.39 | 0.39 | 1 | |
| 9 | 0.36 | 0.3 | 0.32 | 0.38 | 0.33 | 0.38 | 0.36 | 0.36 | 1 |

## Panel Q – 7 Assessors

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | |
| 2 | 0.28 | 1 | | | | | |
| 3 | 0.23 | 0.19 | 1 | | | | |
| 4 | 0.25 | 0.4 | 0.37 | 1 | | | |
| 5 | 0.11 | 0.24 | 0.24 | 0.36 | 1 | | |
| 6 | 0.21 | 0.21 | 0.26 | 0.43 | 0.16 | 1 | |
| 7 | 0.23 | 0.2 | 0.39 | 0.32 | 0.12 | 0.36 | 1 |

## Panel R – 8 Assessors

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | |
| 2 | 0.72 | 1 | | | | | | |
| 3 | 0.75 | 0.74 | 1 | | | | | |
| 4 | 0.8 | 0.83 | 0.82 | 1 | | | | |
| 5 | 0.79 | 0.74 | 0.83 | 0.84 | 1 | | | |
| 6 | 0.61 | 0.65 | 0.65 | 0.65 | 0.69 | 1 | | |
| 7 | 0.61 | 0.73 | 0.72 | 0.66 | 0.82 | 0.64 | 1 | |
| 8 | 0.74 | 0.73 | 0.84 | 0.78 | 0.84 | 0.58 | 0.69 | 1 |

## Panel T – 18 Assessors

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | | | | | | | | | |
| 2 | 0.54 | 1 | | | | | | | | | | | | | | |
| 3 | 0.45 | 0.41 | 1 | | | | | | | | | | | | | |
| 4 | 0.48 | 0.51 | 0.42 | 1 | | | | | | | | | | | | |
| 5 | 0.62 | 0.56 | 0.39 | 0.64 | 1 | | | | | | | | | | | |
| 6 | 0.56 | 0.49 | 0.32 | 0.49 | 0.59 | 1 | | | | | | | | | | |
| 7 | 0.62 | 0.5 | 0.36 | 0.52 | 0.66 | 0.84 | 1 | | | | | | | | | |
| 8 | 0.59 | 0.52 | 0.44 | 0.47 | 0.47 | 0.58 | 0.62 | 1 | | | | | | | | |
| 9 | 0.55 | 0.44 | 0.47 | 0.48 | 0.55 | 0.7 | 0.65 | 0.58 | 1 | | | | | | | |
| 10 | 0.52 | 0.42 | 0.52 | 0.43 | 0.45 | 0.44 | 0.55 | 0.65 | 0.49 | 1 | | | | | | |
| 11 | 0.6 | 0.52 | 0.45 | 0.5 | 0.55 | 0.57 | 0.62 | 0.64 | 0.58 | 0.58 | 1 | | | | | |
| 12 | 0.57 | 0.54 | 0.52 | 0.45 | 0.52 | 0.62 | 0.61 | 0.62 | 0.58 | 0.53 | 0.58 | 1 | | | | |
| 13 | 0.6 | 0.47 | 0.28 | 0.57 | 0.65 | 0.77 | 0.81 | 0.55 | 0.61 | 0.49 | 0.5 | 0.53 | 1 | | | |
| 14 | 0.48 | 0.33 | 0.34 | 0.45 | 0.55 | 0.52 | 0.63 | 0.53 | 0.36 | 0.49 | 0.47 | 0.46 | 0.57 | 1 | | |
| 15 | 0.61 | 0.49 | 0.3 | 0.63 | 0.61 | 0.63 | 0.71 | 0.63 | 0.56 | 0.53 | 0.55 | 0.56 | 0.76 | 0.62 | 1 | 0.75 |
| 16 | 0.56 | 0.34 | 0.36 | 0.41 | 0.49 | 0.69 | 0.77 | 0.63 | 0.58 | 0.62 | 0.57 | 0.59 | 0.69 | 0.62 | 0.75 | 1 |

## Panel U – 11 Assessors

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | | | | |
| 2 | 0.73 | 1 | | | | | | | | | |
| 3 | 0.81 | 0.86 | 1 | | | | | | | | |
| 4 | 0.85 | 0.62 | 0.72 | 1 | | | | | | | |
| 5 | 0.68 | 0.85 | 0.85 | 0.61 | 1 | | | | | | |
| 6 | 0.8 | 0.8 | 0.88 | 0.68 | 0.83 | 1 | | | | | |
| 7 | 0.52 | 0.6 | 0.67 | 0.37 | 0.58 | 0.72 | 1 | | | | |
| 8 | 0.8 | 0.61 | 0.72 | 0.7 | 0.58 | 0.63 | 0.41 | 1 | | | |
| 9 | 0.71 | 0.62 | 0.75 | 0.65 | 0.59 | 0.83 | 0.63 | 0.59 | 1 | | |
| 10 | 0.54 | 0.68 | 0.72 | 0.41 | 0.72 | 0.73 | 0.62 | 0.53 | 0.62 | 1 | |
| 11 | 0.76 | 0.78 | 0.88 | 0.69 | 0.72 | 0.84 | 0.76 | 0.61 | 0.73 | 0.62 | 1 |

**Panel V – 8 Assessors**

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | |
| 2 | 0.40 | 1 | | | | | | |
| 3 | 0.34 | 0.59 | 1 | | | | | |
| 4 | 0.38 | 0.68 | 0.58 | 1 | | | | |
| 5 | 0.33 | 0.65 | 0.67 | 0.78 | 1 | | | |
| 6 | 0.32 | 0.50 | 0.5 | 0.65 | 0.5 | 1 | | |
| 7 | 0.41 | 0.71 | 0.53 | 0.68 | 0.68 | 0.44 | 1 | |
| 8 | 0.50 | 0.62 | 0.56 | 0.71 | 0.75 | 0.57 | 0.72 | 1 |

**Panel W – 10 Assessors**

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|----|
| 1 | 1 | | | | | | | | | |
| 2 | 0.34 | 1 | | | | | | | | |
| 3 | 0.38 | 0.22 | 1 | | | | | | | |
| 4 | 0.5 | 0.27 | 0.24 | 1 | | | | | | |
| 5 | 0.49 | 0.32 | 0.39 | 0.39 | 1 | | | | | |
| 6 | 0.46 | 0.32 | 0.4 | 0.34 | 0.41 | 1 | | | | |
| 7 | 0.54 | 0.3 | 0.36 | 0.4 | 0.37 | 0.55 | 1 | | | |
| 8 | 0.52 | 0.39 | 0.35 | 0.41 | 0.47 | 0.38 | 0.45 | 1 | | |
| 9 | 0.49 | 0.25 | 0.34 | 0.53 | 0.4 | 0.44 | 0.61 | 0.47 | 1 | |
| 10 | 0.5 | 0.36 | 0.44 | 0.36 | 0.38 | 0.42 | 0.48 | 0.55 | 0.35 | 1 |

**Panel Y – 8 Assessors**

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **1** | 1 | | | | | | | |
| **2** | 0.36 | 1 | | | | | | |
| **3** | 0.66 | 0.28 | 1 | | | | | |
| **4** | 0.39 | 0.25 | 0.39 | 1 | | | | |
| **5** | 0.29 | 0.32 | 0.43 | 0.45 | 1 | | | |
| **6** | 0.39 | 0.41 | 0.56 | 0.3 | 0.26 | 1 | | |
| **7** | 0.37 | 0.28 | 0.41 | 0.27 | 0.29 | 0.48 | 1 | |
| **8** | 0.37 | 0.25 | 0.51 | 0.49 | 0.51 | 0.29 | 0.31 | 1 |

**Panel Z – 9 Assessors**

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | | | | | | | | |
| **2** | 0.25 | 1 | | | | | | | |
| **3** | 0.25 | 0.48 | 1 | | | | | | |
| **4** | 0.27 | 0.52 | 0.35 | 1 | | | | | |
| **5** | 0.22 | 0.53 | 0.45 | 0.57 | 1 | | | | |
| **6** | 0.2 | 0.34 | 0.31 | 0.25 | 0.36 | 1 | | | |
| **7** | 0.34 | 0.54 | 0.41 | 0.43 | 0.47 | 0.38 | 1 | | |
| **8** | 0.22 | 0.58 | 0.53 | 0.4 | 0.46 | 0.34 | 0.47 | 1 | |
| **9** | 0.33 | 0.44 | 0.4 | 0.32 | 0.47 | 0.27 | 0.48 | 0.5 | 1 |

## RV Between Panels – All Attributes

The following tables show the RV coefficients calculated to measure the agreement between panels, based on GPA on the average data using all attributes.

|   | N | P | Q | R | T | U | V | W | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|
|   | 1 |   |   |   |   |   |   |   |   |   |
| N | 0.99 | 1 |   |   |   |   |   |   |   |   |
| P | 0.80 | 0.85 | 1 |   |   |   |   |   |   |   |
| Q | 0.93 | 0.93 | 0.8 | 1 |   |   |   |   |   |   |
| R | 0.99 | 0.99 | 0.84 | 0.95 | 1 |   |   |   |   |   |
| T | 0.96 | 0.96 | 0.85 | 0.96 | 0.98 | 1 |   |   |   |   |
| U | 0.99 | 0.99 | 0.83 | 0.97 | 0.99 | 0.98 | 1 |   |   |   |
| V | 0.67 | 0.69 | 0.72 | 0.83 | 0.73 | 0.8 | 0.74 | 1 |   |   |
| W | 0.96 | 0.95 | 0.80 | 0.97 | 0.97 | 0.95 | 0.98 | 0.80 | 1 |   |
| Z | 0.99 | 0.98 | 0.81 | 0.95 | 0.99 | 0.96 | 0.98 | 0.75 | 0.98 | 1 |

## RV Between Panels – Common Attributes

The following tables show the RV coefficients calculated to measure the agreement between panels, based on GPA on the average data using the 5 common attributes.

|   | N | P | Q | R | T | U | V | W | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|
| N | 1 |   |   |   |   |   |   |   |   |   |
| P | 0.96 | 1 |   |   |   |   |   |   |   |   |
| Q | 0.77 | 0.82 | 1 |   |   |   |   |   |   |   |
| R | 0.94 | 0.86 | 0.75 | 1 |   |   |   |   |   |   |
| T | 0.97 | 0.95 | 0.85 | 0.93 | 1 |   |   |   |   |   |
| U | 0.95 | 0.9 | 0.84 | 0.96 | 0.98 | 1 |   |   |   |   |
| V | 0.99 | 0.96 | 0.82 | 0.96 | 0.99 | 0.97 | 1 |   |   |   |
| W | 0.66 | 0.55 | 0.57 | 0.82 | 0.72 | 0.77 | 0.7 | 1 |   |   |
| Y | 0.95 | 0.92 | 0.77 | 0.96 | 0.97 | 0.94 | 0.97 | 0.79 | 1 |   |
| Z | 0.98 | 0.95 | 0.78 | 0.94 | 0.99 | 0.96 | 0.99 | 0.73 | 0.99 | 1 |

## APPENDIX 7: SOME RESULTS FROM INDSCAL

**Panel N**

|  |  |  | Weights | |
| --- | --- | --- | --- | --- |
| **Assessor** | **Stress** | **RSQ** | **Dim 1** | **Dim 2** |
| 1 | 0.189 | 0.879 | 0.935 | 0.068 |
| 2 | 0.173 | 0.833 | 0.853 | 0.325 |
| 3 | 0.259 | 0.728 | 0.832 | 0.192 |
| 4 | 0.161 | 0.901 | 0.927 | 0.204 |
| 5 | 0.166 | 0.907 | 0.939 | 0.160 |
| 6 | 0.224 | 0.886 | 0.934 | 0.117 |
| 7 | 0.185 | 0.925 | 0.961 | 0.042 |
| 8 | 0.200 | 0.787 | 0.856 | 0.233 |
| 9 | 0.219 | 0.851 | 0.918 | 0.092 |
| 10 | 0.219 | 0.705 | 0.767 | 0.342 |
| 11 | 0.170 | 0.856 | 0.633 | 0.675 |
| 12 | 0.264 | 0.800 | 0.885 | 0.126 |
| 13 | 0.264 | 0.734 | 0.847 | 0.126 |
| 14 | 0.165 | 0.905 | 0.939 | 0.154 |
| 15 | 0.179 | 0.922 | 0.953 | 0.116 |

**Panel N – cont.**

| Assessor | Stress | RSQ | Weights | | |
|---|---|---|---|---|---|
| | | | Dim 1 | Dim 2 | Dim 3 |
| 1 | 0.160 | 0.871 | 0.904 | 0.229 | 0.041 |
| 2 | 0.127 | 0.870 | 0.718 | 0.525 | 0.281 |
| 3 | 0.219 | 0.689 | 0.644 | 0.512 | 0.113 |
| 4 | 0.123 | 0.904 | 0.902 | 0.241 | 0.181 |
| 5 | 0.149 | 0.860 | 0.858 | 0.331 | 0.123 |
| 6 | 0.146 | 0.917 | 0.873 | 0.387 | 0.080 |
| 7 | 0.168 | 0.858 | 0.855 | 0.357 | 0.000 |
| 8 | 0.176 | 0.787 | 0.808 | 0.290 | 0.226 |
| 9 | 0.142 | 0.919 | 0.948 | 0.000 | 0.145 |
| 10 | 0.184 | 0.735 | 0.538 | 0.632 | 0.215 |
| 11 | 0.091 | 0.950 | 0.395 | 0.579 | 0.677 |
| 12 | 0.168 | 0.917 | 0.936 | 0.052 | 0.198 |
| 13 | 0.198 | 0.782 | 0.777 | 0.417 | 0.066 |
| 14 | 0.119 | 0.914 | 0.927 | 0.158 | 0.174 |
| 15 | 0.130 | 0.915 | 0.933 | 0.162 | 0.132 |

**Panel P**

|  | | | Weights | |
|---|---|---|---|---|
| Assessor | Stress | RSQ | Dim 1 | Dim 2 |
| 1 | 0.258 | 0.381 | 0.558 | 0.263 |
| 2 | 0.211 | 0.645 | 0.756 | 0.270 |
| 3 | 0.276 | 0.492 | 0.681 | 0.167 |
| 4 | 0.248 | 0.429 | 0.593 | 0.280 |
| 5 | 0.219 | 0.561 | 0.660 | 0.355 |
| 6 | 0.150 | 0.880 | 0.906 | 0.242 |
| 7 | 0.320 | 0.484 | 0.691 | 0.081 |
| 8 | 0.247 | 0.447 | 0.574 | 0.342 |
| 9 | 0.272 | 0.158 | 0.277 | 0.285 |

|  | | | Weights | | |
|---|---|---|---|---|---|
| Assessor | Stress | RSQ | Dim 1 | Dim 2 | Dim 3 |
| 1 | 0.164 | 0.572 | 0.626 | 0.180 | 0.385 |
| 2 | 0.151 | 0.733 | 0.809 | 0.169 | 0.224 |
| 3 | 0.220 | 0.574 | 0.730 | 0.193 | 0.063 |
| 4 | 0.204 | 0.470 | 0.640 | 0.154 | 0.193 |
| 5 | 0.101 | 0.778 | 0.689 | 0.331 | 0.441 |
| 6 | 0.124 | 0.817 | 0.821 | 0.355 | 0.129 |
| 7 | 0.249 | 0.500 | 0.659 | 0.251 | 0.049 |
| 8 | 0.130 | 0.783 | 0.714 | 0.259 | 0.454 |
| 9 | 0.114 | 0.780 | 0.351 | 0.748 | 0.311 |

**Panel Q**

| Assessor | Stress | RSQ | Weights Dim 1 | Dim 2 |
|---|---|---|---|---|
| 1 | 0.432 | 0.239 | 0.039 | 0.488 |
| 2 | 0.320 | 0.108 | 0.216 | 0.248 |
| 3 | 0.287 | 0.253 | 0.214 | 0.455 |
| 4 | 0.231 | 0.599 | 0.656 | 0.411 |
| 5 | 0.225 | 0.712 | 0.828 | 0.165 |
| 6 | 0.364 | 0.044 | 0.189 | 0.092 |
| 7 | 0.269 | 0.328 | 0.275 | 0.502 |

| Assessor | Stress | RSQ | Weights Dim 1 | Dim 2 | Dim 3 |
|---|---|---|---|---|---|
| 1 | 0.314 | 0.554 | 0.073 | 0.741 | 0.009 |
| 2 | 0.188 | 0.380 | 0.308 | 0.378 | 0.378 |
| 3 | 0.195 | 0.339 | 0.195 | 0.476 | 0.272 |
| 4 | 0.172 | 0.478 | 0.552 | 0.278 | 0.311 |
| 5 | 0.159 | 0.771 | 0.841 | 0.165 | 0.190 |
| 6 | 0.197 | 0.415 | 0.444 | 0.124 | 0.450 |
| 7 | 0.204 | 0.306 | 0.189 | 0.337 | 0.396 |

**Panel R**

|  |  |  | Weights | |
| --- | --- | --- | --- | --- |
| Assessor | Stress | RSQ | Dim 1 | Dim 2 |
| 1 | 0.175 | 0.947 | 0.973 | 0.000 |
| 2 | 0.151 | 0.897 | 0.784 | 0.530 |
| 3 | 0.209 | 0.848 | 0.919 | 0.060 |
| 4 | 0.146 | 0.906 | 0.914 | 0.264 |
| 5 | 0.136 | 0.926 | 0.949 | 0.159 |
| 6 | 0.181 | 0.919 | 0.950 | 0.134 |
| 7 | 0.212 | 0.830 | 0.862 | 0.297 |
| 8 | 0.150 | 0.905 | 0.913 | 0.267 |

|  |  |  | Weights | | |
| --- | --- | --- | --- | --- | --- |
| Assessor | Stress | RSQ | Dim 1 | Dim 2 | Dim 3 |
| 1 | 0.090 | 0.989 | 0.994 | 0.035 | 0.000 |
| 2 | 0.146 | 0.913 | 0.729 | 0.069 | 0.614 |
| 3 | 0.207 | 0.824 | 0.749 | 0.513 | 0.009 |
| 4 | 0.066 | 0.981 | 0.935 | 0.000 | 0.328 |
| 5 | 0.157 | 0.904 | 0.817 | 0.456 | 0.168 |
| 6 | 0.166 | 0.885 | 0.861 | 0.360 | 0.121 |
| 7 | 0.161 | 0.891 | 0.582 | 0.683 | 0.293 |
| 8 | 0.128 | 0.916 | 0.669 | 0.624 | 0.281 |

**Panel T**

| | | | Weights | |
|---|---|---|---|---|
| Assessor | Stress | RSQ | Dim 1 | Dim 2 |
| 1 | 0.183 | 0.807 | 0.841 | 0.317 |
| 2 | 0.144 | 0.905 | 0.932 | 0.190 |
| 3 | 0.290 | 0.373 | 0.419 | 0.444 |
| 4 | 0.244 | 0.815 | 0.895 | 0.120 |
| 5 | 0.216 | 0.780 | 0.810 | 0.353 |
| 6 | 0.154 | 0.827 | 0.736 | 0.534 |
| 7 | 0.176 | 0.820 | 0.858 | 0.290 |
| 8 | 0.253 | 0.633 | 0.743 | 0.284 |
| 9 | 0.214 | 0.683 | 0.581 | 0.588 |
| 10 | 0.295 | 0.663 | 0.770 | 0.266 |
| 11 | 0.168 | 0.808 | 0.695 | 0.570 |
| 12 | 0.277 | 0.442 | 0.541 | 0.387 |
| 13 | 0.200 | 0.906 | 0.930 | 0.202 |
| 14 | 0.201 | 0.819 | 0.888 | 0.173 |
| 15 | 0.196 | 0.864 | 0.925 | 0.094 |
| 16 | 0.160 | 0.878 | 0.893 | 0.283 |

**Panel T – cont.**

| Assessor | Stress | RSQ | Weights | | |
|---|---|---|---|---|---|
| | | | Dim 1 | Dim 2 | Dim 3 |
| 1 | 0.122 | 0.856 | 0.746 | 0.488 | 0.249 |
| 2 | 0.104 | 0.926 | 0.755 | 0.589 | 0.093 |
| 3 | 0.218 | 0.545 | 0.670 | 0.168 | 0.261 |
| 4 | 0.110 | 0.879 | 0.361 | 0.837 | 0.220 |
| 5 | 0.124 | 0.815 | 0.666 | 0.515 | 0.325 |
| 6 | 0.133 | 0.831 | 0.718 | 0.349 | 0.441 |
| 7 | 0.137 | 0.852 | 0.708 | 0.525 | 0.274 |
| 8 | 0.116 | 0.828 | 0.358 | 0.718 | 0.429 |
| 9 | 0.115 | 0.815 | 0.495 | 0.372 | 0.656 |
| 10 | 0.195 | 0.679 | 0.590 | 0.540 | 0.199 |
| 11 | 0.1102 | 0.871 | 0.743 | 0.293 | 0.483 |
| 12 | 0.162 | 0.656 | 0.620 | 0.329 | 0.404 |
| 13 | 0.110 | 0.874 | 0.487 | 0.766 | 0.224 |
| 14 | 0.155 | 0.796 | 0.653 | 0.600 | 0.100 |
| 15 | 0.125 | 0.882 | 0.440 | 0.825 | 0.089 |
| 16 | 0.126 | 0.850 | 0.707 | 0.568 | 0.169 |

**Panel U**

| | | | Weights | |
|---|---|---|---|---|
| Assessor | Stress | RSQ | Dim 1 | Dim 2 |
| 1 | 0.086 | 0.976 | 0.979 | 0.132 |
| 2 | 0.195 | 0.195 | 0.883 | 0.214 |
| 3 | 0.109 | 0.942 | 0.893 | 0.379 |
| 4 | 0.134 | 0.134 | 0.983 | 0.000 |
| 5 | 0.123 | 0.925 | 0.824 | 0.497 |
| 6 | 0.105 | 0.105 | 0.835 | 0.499 |
| 7 | 0.187 | 0.833 | 0.491 | 0.769 |
| 8 | 0.155 | 0.155 | 0.878 | 0.386 |
| 9 | 0.160 | 0.891 | 0.871 | 0.363 |
| 10 | 0.125 | 0.125 | 0.766 | 0.595 |
| 11 | 0.117 | 0.956 | 0.809 | 0.549 |

| | | | Weights | | |
|---|---|---|---|---|---|
| Assessor | Stress | RSQ | Dim 1 | Dim 2 | Dim 3 |
| 1 | 0.086 | 0.962 | 0.930 | 0.245 | 0.192 |
| 2 | 0.117 | 0.934 | 0.623 | 0.042 | 0.737 |
| 3 | 0.093 | 0.957 | 0.658 | 0.484 | 0.538 |
| 4 | 0.079 | 0.988 | 0.986 | 0.130 | 0.000 |
| 5 | 0.071 | 0.973 | 0.462 | 0.509 | 0.707 |
| 6 | 0.106 | 0.945 | 0.610 | 0.741 | 0.154 |
| 7 | 0.176 | 0.924 | 0.142 | 0.951 | 0.012 |
| 8 | 0.135 | 0.925 | 0.726 | 0.516 | 0.362 |
| 9 | 0.168 | 0.882 | 0.737 | 0.567 | 0.132 |
| 10 | 0.139 | 0.921 | 0.514 | 0.792 | 0.173 |
| 11 | 0.095 | 0.974 | 0.525 | 0.827 | 0.116 |

**Panel V**

| Assessor | Stress | RSQ | Weights | |
|---|---|---|---|---|
| | | | Dim 1 | Dim 2 |
| 1 | 0.282 | 0.286 | 0.417 | 0.335 |
| 2 | 0.219 | 0.534 | 0.592 | 0.429 |
| 3 | 0.304 | 0.557 | 0.738 | 0.113 |
| 4 | 0.193 | 0.694 | 0.759 | 0.345 |
| 5 | 0.113 | 0.972 | 0.977 | 0.132 |
| 6 | 0.193 | 0.731 | 0.803 | 0.293 |
| 7 | 0.255 | 0.475 | 0.606 | 0.328 |
| 8 | 0.202 | 0.791 | 0.825 | 0.333 |

| Assessor | Stress | RSQ | Weights | | |
|---|---|---|---|---|---|
| | | | Dim 1 | Dim 2 | Dim 3 |
| 1 | 0.188 | 0.235 | 0.345 | 0.248 | 0.234 |
| 2 | 0.160 | 0.640 | 0.598 | 0.513 | 0.140 |
| 3 | 0.217 | 0.602 | 0.754 | 0.089 | 0.162 |
| 4 | 0.158 | 0.705 | 0.734 | 0.362 | 0.188 |
| 5 | 0.073 | 0.974 | 0.973 | 0.159 | 0.051 |
| 6 | 0.166 | 0.680 | 0.756 | 0.192 | 0.268 |
| 7 | 0.167 | 0.507 | 0.583 | 0.329 | 0.242 |
| 8 | 0.145 | 0.773 | 0.802 | 0.323 | 0.162 |

**Panel W**

| | | | Weights | |
|---|---|---|---|---|
| Assessor | Stress | RSQ | Dim 1 | Dim 2 |
| 1 | 0.262 | 0.424 | 0.592 | 0.272 |
| 2 | 0.139 | 0.833 | 0.485 | 0.773 |
| 3 | 0.290 | 0.221 | 0.326 | 0.338 |
| 4 | 0.247 | 0.625 | 0.770 | 0.178 |
| 5 | 0.275 | 0.451 | 0.643 | 0.192 |
| 6 | 0.291 | 0.427 | 0.637 | 0.145 |
| 7 | 0.308 | 0.408 | 0.618 | 0.162 |
| 8 | 0.261 | 0.509 | 0.685 | 0.198 |
| 9 | 0.259 | 0.572 | 0.742 | 0.146 |
| 10 | 0.324 | 0.295 | 0.528 | 0.130 |

| | | | Weights | | |
|---|---|---|---|---|---|
| Assessor | Stress | RSQ | Dim 1 | Dim 2 | Dim 3 |
| 1 | 0.182 | 0.514 | 0.407 | 0.554 | 0.202 |
| 2 | 0.145 | 0.801 | 0.370 | 0.183 | 0.794 |
| 3 | 0.166 | 0.458 | 0.338 | 0.349 | 0.471 |
| 4 | 0.157 | 0.676 | 0.653 | 0.483 | 0.125 |
| 5 | 0.112 | 0.837 | 0.594 | 0.679 | 0.152 |
| 6 | 0.179 | 0.684 | 0.785 | 0.231 | 0.121 |
| 7 | 0.201 | 0.693 | 0.806 | 0.123 | 0.167 |
| 8 | 0.218 | 0.493 | 0.554 | 0.412 | 0.128 |
| 9 | 0.190 | 0.627 | 0.720 | 0.312 | 0.106 |
| 10 | 0.210 | 0.568 | 0.711 | 0.224 | 0.110 |

**Panel Y**

| Assessor | Stress | RSQ | Weights | |
|---|---|---|---|---|
| | | | **Dim 1** | **Dim 2** |
| 1 | 0.289 | 0.498 | 0.692 | 0.138 |
| 2 | 0.196 | 0.572 | 0.480 | 0.580 |
| 3 | 0.111 | 0.945 | 0.970 | 0.067 |
| 4 | 0.227 | 0.613 | 0.733 | 0.275 |
| 5 | 0.291 | 0.199 | 0.389 | 0.218 |
| 6 | 0.192 | 0.842 | 0.911 | 0.116 |
| 7 | 0.314 | 0.436 | 0.653 | 0.993 |
| 8 | 0.254 | 0.471 | 0.639 | 0.249 |

| Assessor | Stress | RSQ | Weights | | |
|---|---|---|---|---|---|
| | | | **Dim 1** | **Dim 2** | **Dim 3** |
| 1 | 0.262 | 0.544 | 0.726 | 0.123 | 0.047 |
| 2 | 0.148 | 0.661 | 0.496 | 0.625 | 0.154 |
| 3 | 0.072 | 0.976 | 0.987 | 0.043 | 0.032 |
| 4 | 0.145 | 0.703 | 0.770 | 0.248 | 0.222 |
| 5 | 0.120 | 0.660 | 0.559 | 0.333 | 0.487 |
| 6 | 0.157 | 0.806 | 0.881 | 0.086 | 0.153 |
| 7 | 0.197 | 0.630 | 0.703 | 0.081 | 0.359 |
| 8 | 0.166 | 0.584 | 0.691 | 0.255 | 0.203 |

**Panel Z**

| Assessor | Stress | RSQ | Weights Dim 1 | Dim 2 |
|---|---|---|---|---|
| | | | **Weights** | |
| 1 | 0.234 | 0.630 | 0.625 | 0.488 |
| 2 | 0.338 | 0.500 | 0.701 | 0.096 |
| 3 | 0.330 | 0.524 | 0.718 | 0.090 |
| 4 | 0.331 | 0.342 | 0.559 | 0.174 |
| 5 | 0.324 | 0.305 | 0.506 | 0.222 |
| 6 | 0.252 | 0.554 | 0.426 | 0.610 |
| 7 | 0.316 | 0.525 | 0.705 | 0.168 |
| 8 | 0.230 | 0.716 | 0.813 | 0.238 |
| 9 | 0.185 | 0.755 | 0.648 | 0.579 |

| Assessor | Stress | RSQ | Weights Dim 1 | Dim 2 | Dim 3 |
|---|---|---|---|---|---|
| | | | **Weights** | | |
| 1 | 0.206 | 0.458 | 0.543 | 0.375 | 0.149 |
| 2 | 0.206 | 0.574 | 0.711 | 0.065 | 0.254 |
| 3 | 0.158 | 0.728 | 0.710 | 0.053 | 0.470 |
| 4 | 0.229 | 0.345 | 0.502 | 0.127 | 0.278 |
| 5 | 0.164 | 0.540 | 0.533 | 0.230 | 0.450 |
| 6 | 0.143 | 0.810 | 0.497 | 0.735 | 0.149 |
| 7 | 0.211 | 0.561 | 0.716 | 0.134 | 0.174 |
| 8 | 0.162 | 0.729 | 0.803 | 0.251 | 0.147 |
| 9 | 0.121 | 0.815 | 0.622 | 0.614 | 0.226 |

**2-Dimensional Plots**

Panel N - Consensus INDSCAL Plot - Tomato Soup



Panel P - Consensus INDSCAL Plot - Tomato Soup

# Panel Q - Consensus INDSCAL Plot - Tomato Soup



# Panel R - Consensus INDSCAL Plot - Tomato Soup

# Panel T - Consensus INDSCAL Plot - Tomato Soup



# Panel U - Consensus INDSCAL Plot - Tomato Soup

## Panel V - Consensus INDSCAL Plot - Tomato Soup



## Panel W - Consensus INDSCAL Plot - Tomato Soup

## Panel Y - Consensus INDSCAL Plot - Tomato Soup

F1

C1F1

F2

Dim 2

C1

Base

C1F2

-1    0    1

Dim 1

## Panel Z - Consensus INDSCAL Plot - Tomato Soup

Base

C1

Dim 2

C1F1

F1

C1F2

F2

-2    -1    0    1    2

Dim 1