Sensory Evaluation – Analysis of the Reliability of Expert Panelist Testing

Precision Scientific Software Inc. (PSS) has developed a program (Attribute R&R) to analyze the repeatability and reproducibility (R&R) of subjective evaluations, such as sensory evaluations, conducted by expert panelists.

The program looks at two factors of reliability:
- Repeatability is the level of agreement between ratings given by an individual panelist when doing multiple evaluations of the same sample (does the panelist consistently give the same rating each time)
- Reproducibility is the level of agreement between ratings given by multiple panelists evaluating the same sample (how consistent are the ratings given by all the panelists as a group)

The importance of these two factors is that they both contribute to the variability inherent in the measurement.

Whenever you make a measurement you have to understand how much of the variability in your data comes from the products you're measuring and how much comes from the method you're using to make the measurement. And you want to be able to minimize the amount of variability that comes from your measurement system.

If the variability due to the measurement system (i.e., the repeatability and reproducibility) is large then it "swamps" the true variability or the true difference between the products. The danger of this situation is that you may make business decisions based on the variability inherent in your measurements rather than on the true differences between products.

PSS also created a "sister" program (Gauge R&R) to quantify the repeatability and reproducibility of objective laboratory evaluations such as the chemical analysis of the levels of CBD's in cannabis products.

Each of these programs applies the same considerations to their respective fields:
- Rigorous statistical analysis of the data. This avoids the possibilities of bias and inconsistency that could be introduced if the data is analyzed based on the intuition of the analyst
- Maximum automation to relieve the analyst of the most time-consuming and error-prone tasks. The programs:
  - Fully automate the tedious tasks of giving random codes to the samples and randomizing the order of presentation of the samples to the panelists
  - Generate questionnaires for each panelist in the same order as the samples will be presented
  - Capture the data directly from the questionnaires to avoid transcription error
  - "De-randomize" the data to ensure that each value is correctly assigned to the appropriate data element in the statistical table
- Automatically calculate all relevant statistical values
- Prepare a summary report of top-line results suitable for presentation to management

Periodically conducting this type of R&R Study enables you to:
- Evaluate the level of proficiency of your panelists or your laboratory.
  - Repeatability and reproducibility are key factors in the Estimation of Measurement Uncertainty which is a requirement of accreditation to the ISO 17025 Lab Proficiency International Standard
- Determine how much of the total variability is contributed by the measurement system
  - Assess whether the measurement methodology is adequate to meet the business requirements
  - If the level of variability is small this means you can reliably detect actual differences between products
  - But if the level of variability is large then this may obscure the true differences between products or it may incorrectly indicate that two products are different even though in reality they are equivalent
  - Not knowing whether the measurement system is adequate may result in making incorrect business decisions due to being unaware of the limits of the reliability of your measurements
- Identify the need for remedial action
  - If the measurement system is found to be inadequate to meet the business requirements (for instance if the variability is too large compared to the degree of difference between products you are trying to detect) then you can take steps to reduce the variability by re-training or re-calibrating the panelists or you can show justification to purchase more precise instrumentation for the laboratory