# Consumer Product Evaluation, Principles and Practice.

**Sitaram Dixit**
**Senior Manager Fragrance Applications, International Flavours and Fragrances (India) Limited.**

The advances that human knowledge has made is not by the quirk of chance but largely due to the use of experimental and statistical methods. Consumer products once were marketed without any justification of claims. Competition, an increasing awareness, literacy, better education, overall economic growth of the consumers, and Government legislation have forced manufacturers to confirm and fulfil mandatory specifications in modern times. Consumers like and needs are most important to be fulfilled to be successful in the consumer product industry. This paper intended to a non-statistical non-expert gives an over view of consumer product evaluation, and illustrates the enumerated principles with identifiable live examples.

We acquire knowledge from different sources. In early ages, acquiring knowledge was considered a matter of pure chance, the origin of the acquired knowledge practically unknown. The advances that human knowledge has made today is not by the quirk of chance but largely due to the adoption of tried and tested methods followed systematically and according to a plan. Methods following a plan and systematically used in broadening knowledge are known as scientific methods.

During the infancy of our industry, products were launched and marketed without any justification of the claims that were made by the manufacturers. In modern times, manufacturers have to contend with tough competition. There is also an increasing awareness among consumers about various products available in the market place, due to better education, literacy, and overall economic growth of the masses. Moreover, Government legislation and regulatory authorities have forced manufacturers to confirm and fulfil the minimum required specifications. If one has to ensure success in the market, one should know more about not only ones product performance but also that of the competition.

Product evaluation is a science dealing with subjects that is neither private nor confidential. It is not designed for the specially initiated group but easily understood by all. However, it does follow the rules of any science, namely observation, hypothesis, measurement, explanation, and conclusion. Evaluation means to determine or to examine and judge. It suggests an attempt to determine either the relative or the intrinsic worth of something in terms other than monetary. Product evaluation thus can be defined as the assessment of one or more of the attributes, characteristics or qualities of a specific product.

Some typical examples are

- To evaluate whether brand A of face cream is greasier than brand B, under normal conditions of use.
- To evaluate the efficacy of a deodorant roll-on
- To evaluate the anti-dandruff efficacy of an experimental shampoo having a natural ingredient viz. a viz., a synthetic active based anti-dandruff shampoo.

Practical techniques used for product evaluation are sometimes simple, and sometimes a little complicated. They also have to fulfil the following as its main functions.

1. **New product development:** Evaluation helps in the development of new products. It helps to create end products that have the required product attributes and properties.

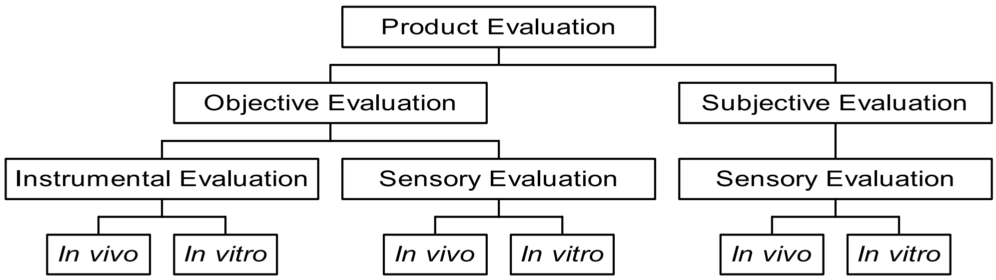
2. **Formula improvements in existing products:** Product evaluation is also used to determine change in attributes by use of new or different raw materials in existing products. When any new or additional ingredients are used for product improvements, changes are made to reduce formulation cost, there is change in supplier or replacement due to raw material shortage; product evaluation plays a very important role in final product quality.

3. **Comparing performance between competing products:** In this industry, every manufacturer needs to know as to how his product performs against competitive products available in the market. Product evaluation helps one understand the quality of competing products and identify attributes required to improve one's product to ensure success in the market place.

4. **Product claim validation:** In today's modern world valid scientific documentation are necessary to be submitted if the manufacturer has to advertise specific claims for products marketed. Product evaluation helps one substantiate various claims made by the marketing department.

**PRODUCT EVALUATION TECHNIQUES AND ITS TYPES**

Product evaluation techniques can be divided into various types. This division helps us to set the limit up to which the product evaluation technique would confirm and classify our requirements.




**SUBJECTIVE EVALUATION:**

Subjective evaluation concerns procedures that involve sensory test methods, the results of which are obtained from a group of panellist, wherein each panellist gives results that are largely influenced by personal preferences. Subjective evaluation techniques does not generally require well trained persons and can be carried out by untrained personnel under the supervision of a trained leader.

Subjective evaluation in the final analysis is very closely related to the likes and dislikes of the product by which the consumer decides either to purchase it or not. Subjective tests like consumer research and consumer evaluation tests are designed for this purpose to understand consumer preferences. The number of available participants in any laboratory is usually small and is also not representative of the consumer population base or target group. This problem is all the more profound when product design and development work is carried out in one place and marketing is to be done in a totally different area. Likes and dislikes are not universal. Requirements differ from place to place and country to country. However well one knows the tastes of the consumers in one place, it is very difficult to predict consumer product preference based on the result of consumer tastes of a different place.

**OBJECTIVE EVALUATION:**

It concerns those procedures that are carried out and supervised by trained persons. The evaluators either by use of sensory or instrumental means generate results that are quantitative and reproducible in nature. Objective test methods can either be sensory or instrumental sensory. It can be further subdivided into psychophysical and psychometric evaluation

**1. PSYCHOPHYSICAL EVALUATION:**

Psychophysical test methods make use of psychological judgements where decisions are made by direct physical measurements. E.g., size of the shampoo bottles, length and breath of a detergent bar, etc. In both these examples the property to be investigated can be evaluated by an objective sensory procedure, namely ranking a group of sample based on the size visually or by an objective instrumental method by using a measuring scale to give the length and breath in centimetres.

**2. PSYCHOMETRIC EVALUATION:**

Psychometrics is the psychological measurement of parameters for which exists no direct physical measurements. E.g., Oiliness imparted by a face cream. Psychometrics can sometimes also extend into subjective sensory evaluations. One has to be cautious while defining objective sensory evaluation. The initiator of the evaluation tests has to clearly understand and should have no doubts about the questions for which answers are to be found. The limitation of extrapolation of the obtained data to answer other unanswered questions should also be clearly understood.

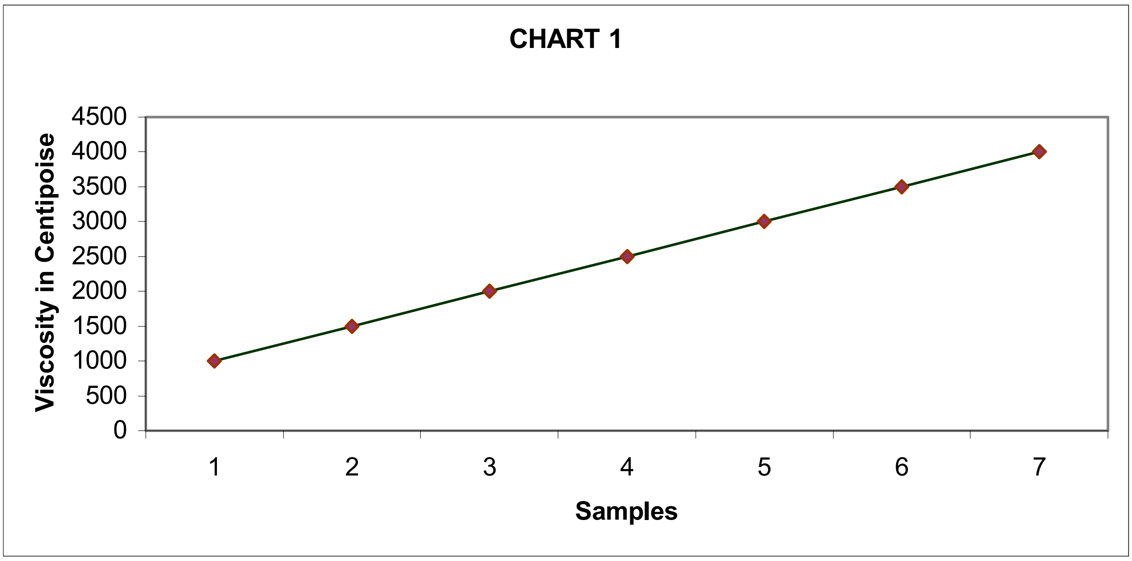
Different product evaluation tests will provide different results when the same product is evaluated by use of a different set of questions. When the questions asked are different or when the questions are not related linearly, the results obtained would differ substantially and widely. Let us consider a very simple example as an illustration of how the questions asked by the initiator can drastically change the results of an evaluation test carried out.

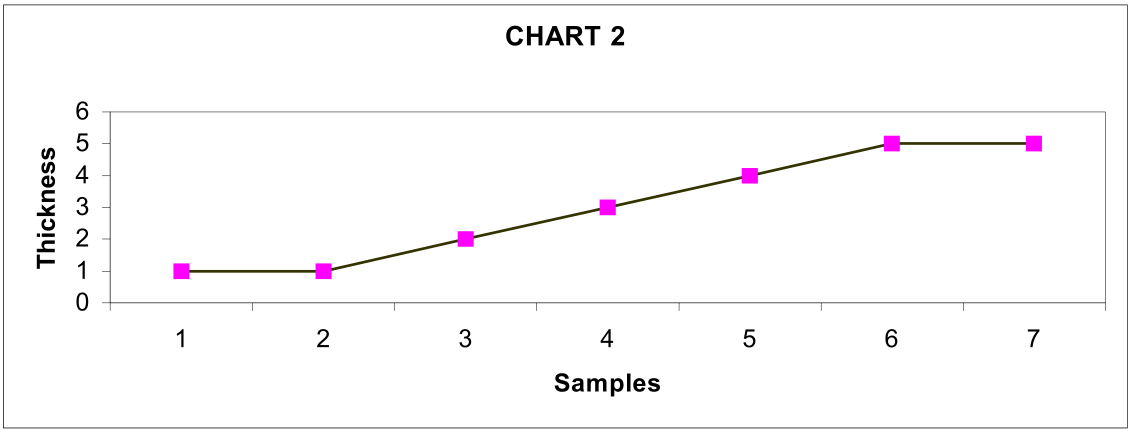
Ms. Clean is a regular user of liquid hand wash soap, who prefers the viscosity of the product to be around 2500 centipoise (cps) during use. She strongly believes that liquid soap should be neither too light nor too thick for superior performance and ease in use. Supposing we give Ms. Clean, seven different samples of liquid hand wash soap one sample having a viscosity of 1000 cps and gradually increasing by 500 cps upto 4000 cps.

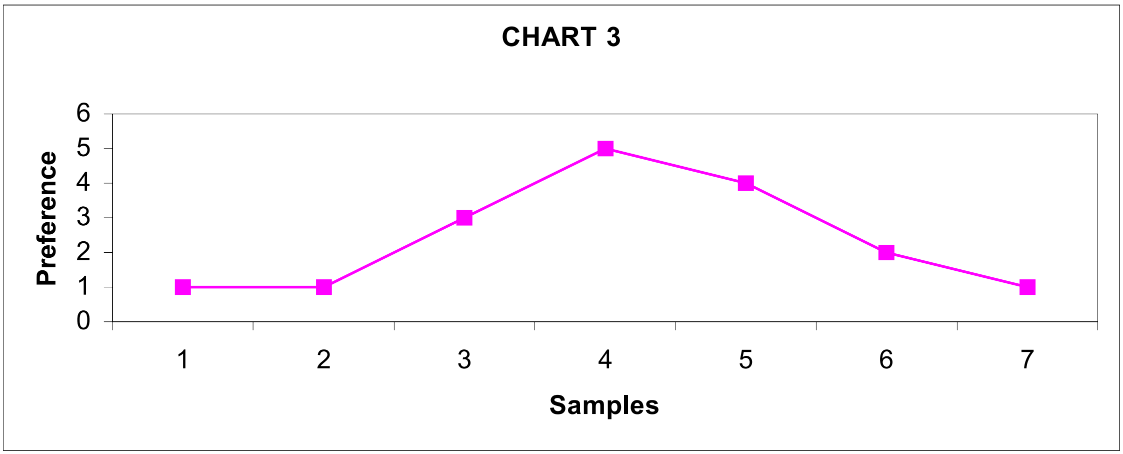
In our first test, suppose we ask Ms. Clean to evaluate the viscosity of the liquid soap given. The obvious method followed would be to make use of a viscometer and determine precisely the viscosity of the samples provided to her. Thus the test would be an objective instrumental one carried out in vitro and the test results when graphically represented will look as shown in CHART 1.

In the second test suppose we ask Ms. Clean to evaluate the liquid soap for thickness on a five point scale **(0 is very thin to 5 very thick)**, the result graphically would probably look like CHART 2. This is typical case of an objective sensory test carried out in vivo. The response is not linear as in the first because there is difficulty in detecting differences in both low and high viscosity levels, than in the intermediate ranges.

Finally supposing the question asked to Ms. Clean is to give her preference based on thickness for a hand wash liquid soap using a similar five point scale **(0 not liked to 5 extremely liked)**. The results are most likely depicted graphically as CHART 3, with the bliss point or the peak of acceptability at the top of the graph.






The above tests clearly shows that questions asked to the respondents plays a major role in understanding the various parameters required in a consumer product. This very important aspect of evaluation should always be remembered when carrying out any objective product evaluation. Another example would be the examination for smoothness of application of a skin cream. The result can range from very smooth to course and the extreme smoothness of the cream may not be actually appreciated for being too slick and oily.

**SAMPLING AND SAMPLING DESIGN PRECAUTIONS**

A sample is not studied for its own sake. The basic objective of sampling and its study is to draw inference about the population. In other words, sampling is a tool that helps to know the characteristics of the universe or the population by examining only a small part of it.

A sample selection must be planned carefully and executed correctly, otherwise the results obtained may be inaccurate and misleading. It is obvious that unless the samples selected are representative, the results obtained by the study will be absolutely useless without any value attached to them. However, one should remember that accidental biases could creep in the results due to its style of presentation, even if utmost care has been taken in selecting the sample.

Sensory evaluation test results of a particular parameter or attribute can be easily influenced by the information received by our brain by use of our other senses. For example a cosmetic floral fragrance when put in a green coloured shampoo may be erroneously referred to as an herbal shampoo fragrance. Samples labelled alphabetically or numerically can be incorrectly ranked because “A” or “1” may be considered to as a superior than that of succeeding letters or numbers. The order in which the samples are presented to the evaluators may also bias the end results inadvertently. The following points when observed while carrying out sensory evaluation tests can largely help in eliminating the biases observed in the evaluation procedure.

- Respondents who are able to identify the samples are not to be used as evaluators and should be eliminated.
- Two or three digit random numerical code, multiple alphabetical code, or a combination of numerical and alphabetical codes should be used.
- Samples selected for evaluation should be packed in identical packaging material
- All samples selected should have the same temperature.
- All samples used for evaluation should have the same colour. Colour should never be used for identification. In case the samples are of different colour, the evaluator should be necessarily blindfolded to rule out any colour bias.
- Samples of the same size and shape should only be presented to the respondents during evaluation.
- Samples should be presented in a random order and not in a fixed sequence.
- Evaluating a large number of samples continuously leads to fatigue. This all the more important when odour evaluation of samples are carried out. The number of sample should be staggered so as to rule out evaluation fatigue.
- When fragrance stability or preference is investigated in the end product, usage of the same quality of raw material to make sample is necessary. This helps in the required parameter only being highlighted for evaluation all other attributes being constant.
- The sample should preferably have only one variable with all other parameters constant. Too large a number of variables in a sample would only lead to confusion and erroneous result.

## STATISTICAL METHODS IN OBJECTIVE EVALUATION

Classical or parametric statistical methods and Non-parametric statistical methods, both are used to draw conclusions in objective evaluation.

## PARAMETRIC STATISTICAL METHODS.

Statistical parametric classical tests are the most appropriate to be chosen, and so should be used in determining the truth of the probability statement, when the necessary test criteria's are met. In order to apply Parametric statistical methods the data used must confirm the following basic requirements.

- The sample data collected for analysis should belong to normally distributed population.
- The sample population should have same variance or should have a known ratio of variance. In short, the data must be a linear combination of effects.
- Each datum must be independent of the other. The selection of any one instance from the sample population for inclusion in the sample should not influence or bias the chances of the other case for inclusion. The score obtained in one instance must not influence or bias the score of other.

When the type population distribution is known, the tests of significance are made on assumptions. Sometimes a test of significance might be based on the assumption that the sample value were drawn from a normally distributed universe or that the two samples were drawn from universes having the same variance. The testing procedure also assumes that the unknown values of the parameters about which statistical inferences were to be made could be estimated from statistic obtained from random samples. This approach to inferential statistics is called parametric methods, since the concern here is with the value of the parameter.

## NON-PARAMETRIC STATISTICAL METHODS.

There are many situations where it is not possible to make rigid assumptions about the shape of the population from which samples are drawn. This limitation led to the development of a group of alternate techniques known as non-parametric or distribution free methods. A non-parametric method may be defined as a statistical test in which no hypothesis is made about specific values or parameters. Distribution free tests may be defined as methods for testing a hypothesis that does not depend on an assumption concerning the form of the underlying distribution. Thus when the data available do not fulfil the mandatory test criteria, non-parametric tests are applied. Although non-parametric test conclusions are generally considered to be less powerful than classical parametric test methods, the results are nevertheless reliable and do help us draw conclusions to take decisions.

***ADVANTAGES AND DISADVANTAGES OF NON-PARAMETRIC TEST METHODS.***

- The accuracy of the probability statements does not depend on the shape of the sample population.
- Sometimes non-parametric tests assume the shape of two or more population distribution or otherwise assume symmetrical population distribution. It can also assume an underlying continuous distribution similar to parametric tests.
- In most non-parametric statistical tests the probability statements obtained are exact probabilities or very good approximations, no matter what shape of the population distribution from which the random samples are drawn.
- Non-parametric statistical tests are very useful when sample size is small and the nature of the population distribution is not very well known.
- Sometimes research is only able to say that one samples has more or less of the specific attribute than another without being able to quantify the data as to how much more or less, for example when the data's are given in ranks or categorised as plus or minus. Such data having the strength of ranks are more easily dealt by using non-parametric test methods. If parametric test methods are to be used in such cases then risky and unrealistic assumptions may have to be made.
- Data that are measured in a nominal scale and classifying in nature cannot be tested by use of classical parametric test methods. Non-parametric methods are useful to deal with them.
- Non-parametric test methods are very easy to understand and put in practice than parametric test methods.
- If all the mandatory test criteria requirement of data necessary for parametric statistical test is fulfilled and also if the results required should have a lot of weightage, the non-parametric statistical tests are not useful and when applied is only an utter waste of time and effort.
- Testing interaction in the analysis of variance model cannot be carried out by use of non-parametric statistical test methods unless specific assumptions are made about additivity.

**OTHER REQUIREMENTS**

Along with the collection of data, sampling techniques, presentation skills, other considerations like the quality of the assessors and type of rating scales are also equally important and should be looked into before starting any objective evaluation tests. All evaluation tests conducted and the results so produced have to be necessarily reproducible and quantifiable. This is possible only if the assessors posses the following four major qualities.

1. ***Reproducibility of judgement:*** The assessor doing the evaluation must be able to repeat their judgements so that the results obtained by their assessments are reproducible. The ability of the assessor to be consistent need to be checked by carrying out reliability tests, with the same, but differently coded samples over a period of time before their selection in the evaluator panel.

2. ***Objective Nature:*** Assessors selected should not be influenced by any other attribute other than that we require them to assess and give their valued responses. Assessors selected in the panel should be tested on their objective approach by deliberately biasing them and seeing if their judgements are influenced or are different when tests are conducted in an unbiased situation.

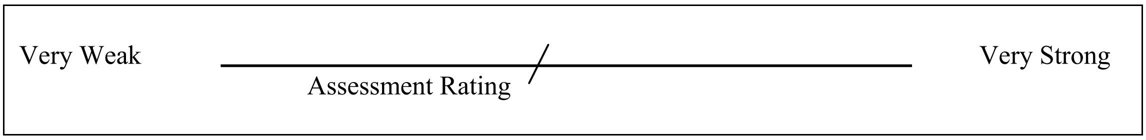
3. ***Inherent interest in being an assessor:*** Assessors selected should be interested and willing to take part in the evaluation tests. If assessors are not interested than their judgement can be erroneous. Interest in the respondents should be maintained, by explaining evaluation process and the value of their judgements in product development. Regular responses of results should be provided to the assessors so that their interest in the evaluation process is retained until completion. Incentives either in cash or kind should be a offered to stimulate competitiveness amongst the group as lack of interest and willingness to participate will only make the tests less sensitive and meaning less.

4. ***Assessors as good communicators:*** Good communication between the questioner or the investigator and the assessor is a very important aspect that is essential for quality objective testing. The respondents should not only be able to understand the questions asked but also be able to communicate back his judgement in a precise manner without any ambiguity. This is very important especially when the attribute being investigated is of a complex nature. E.g., odour

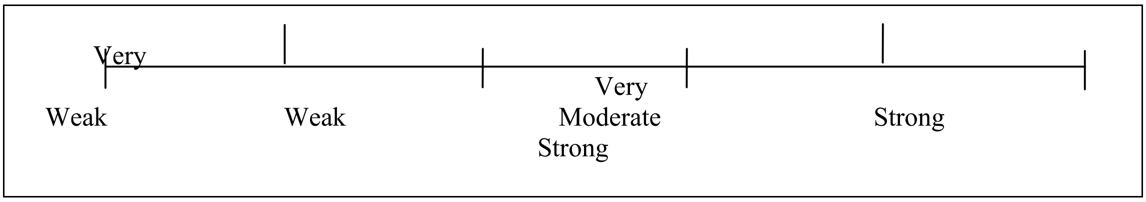
profile of a fragrance in a product, greasiness of a skin cream or lotion, etc. The assessors should easily understand descriptors used in the questionnaires, designed for carrying out the evaluation process. In case there are doubts existing in their minds they should be clearly explained to, so that both the investigator and the assessors understand the same common meaning to the word described. A trained panel of assessors with a well planned and designed sensory evaluation technique can provide a reliable, consistent, accurate and valuable insight about the product to be assessed and help the development chemist during all stages of product development.

**RATING SCALES:**

Rating scales provide the general framework against which an attribute can evaluated. The main types of rating scales used, are the graphical, verbal, and numerical scales. Graphical scale is a very simple rating scale that is like a straight line the two ends of which is marked with the extremes of the attribute to be investigated. The assessor marks across the line to indicate the choice on the measure of the attribute being assessed. For e.g., Strength of the fragrance in toilet soaps.




If the scale is broken down to segments, it will be as below.




When graphical scales are individually labelled then it becomes a verbal scale.

☐ Very Weak ☐ Weak ☐ Moderate ☐ Strong ☐ Very Strong

If the rating scale is in a series of numbers instead of statements, it is considered to be numerical

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

However, the preferred rating scale is the one that is a combination of both numerical and verbal scales. For e.g.,

Property tested: Odour Intensity of a fragranced soap sample at 45 degree centigrade after 1 month in comparison with sample kept at 25 degree centigrade.

Rating scale used is as given below.

| 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Unchanged | Very Slightly Weakened | Slightly Weakened | Markedly Weakened | Strongly Weakened | No Perfume |

Sometimes rating is according to Scale of standards. In this case, actual samples that represent the distinct scale intervals are provided to the assessor to rate the sample. If it is not possible to provide standards across the entire scale, then only some of them called markers are provided to indicate intermittent or intermediate scales on the graphical, verbal or numerical scales, This is termed as a Partial standard scale.

Larger the number of segments in the rating scale better the sensitivity of the evaluation process. However, if the segmentation is too large or too small, then one can encounter difficulty to discriminate between samples, the optimum rating scale segment being between 5 to 10. Other important aspects that one should remember when setting up the rating scale is that scales sometimes can also skew the data making the distribution non - normal . Products can become clustered around one end of the scale and are skewed because there is no negative descriptor to extend the scale. For e.g., Nature of a Lip cream

| | | | |
|---|---|---|---|
| □ Not Oily | □ Very slightly Oily | □ Slightly Oily | □ Moderately Oily |

**GENERAL STEPS FOR OBJECTIVE EVALUATION**

Evaluation tests can become complex and this can confuse the evaluators, making them lose track and control of the tests, leading to wrong and meaning less results and conclusions. However, when the entire evaluation procedure set up is well planned in a serialised protocol with a standard operating procedure, it becomes simple without any complexity or ambiguity. The following are the main steps to be followed in order to avoid unwanted problems that can otherwise occur.

1. **Setting up the Objective:** The exact objective of the process should be clearly defined before starting any evaluation exercise. The various parameters and precise conditions under which the information is to be obtained should be confirmed. E.g., To evaluate the deodorant efficacy of an experimental roll-on deodorant viz., a viz., the leading brand "Rx" available in the market under normal conditions of use.

2. **Setting the outlines of the method to use:** The method to be used for evaluating should be considered in details. Various factors like sensitivity required, complexity of the evaluation process, the cost that can be incurred, the time available at our disposal, are all important and have to be weighed before deciding the method to use, for obtaining data. It is also important to consider the method to be used to test the validity of the results obtained and the manner the results is to be expressed for easy understanding. For example, during deodorant testing one usually do not carry out major clinical tests on the product especially in the initial stages of product development, Nevertheless screening tests are carried out to indicate whether the development is moving in the correct direction or not. Evaluating body odour by sniffing (sniff test) the under arm with and without use of the deodorant is the most sensitive and reproducible method. The tests can be direct or indirect. In an indirect test, the odour from the test site is trapped in an odour trapping and collecting device and then evaluated by the subject or by a panel of trained evaluators. In a direct sniff test, the test site itself is sniffed to obtain the rating. The sniffing is done by the subject either himself or herself or assessed by an evaluator panel of trained sniffers.

3. **Setting up a detailed protocol for operations:** A step by step set of instructions or standard operating procedures are written down in sufficient details, so that any person is in a position to repeat the test in an identical manner to get a precise, accurate and reproducible result

4. **Validating the results obtained after the evaluation:** All results obtained by evaluation should be validated statistically before drawing any conclusions.

5. **Conclusions:** The statistically validated test result is used to draw conclusions and find answers to the defined objective.

All the steps described above are very easy to understand when practically applied in different examples of objective evaluations as encountered in our industry. The examples given below will make this clear.

**PAIRED COMPARISON TEST**

In a paired comparison test one sample is chosen between two alternative set of samples. This being a very simple process it is extensively used in objective evaluation tests.

**Example:**

Your competition has introduced a face cream in the market. Indications from the consumer are that the new cream is greasier than your popular brand in the market. The marketing team wants to know whether this is true to work out the marketing strategy. The results are required within two hours before the start of the monthly meeting with the directors in the afternoon.

**Objective:**

To evaluate whether the competition sample "X" is more greasier than "Y" your brand of face cream under normal conditions of use.

**Method to use:**

- Time available at disposal is only two hours for conducting the evaluation tests
- The samples "X" and "Y" should be tested on the same persons as subjects for comparative tests.
- The test is of sensitive nature and should be carried out in vivo, as instrumental test methods are not suitable.
- The test suitable would be a paired comparison test using at least 20 respondents. Due to time constraints and non-availability of respondents at short notice, only 14 persons can be used.
- Results are required within a short time. It can be quickly analysed statistically by use of Binomial test.

**Protocol and Standard operating procedure:**

- 14 female volunteers are asked to wash and clean their face with a standard soap. The face is dried with a clean towel or tissue.
- The respondent is given the first cream. The cream can either be 'X" or "Y"
- The respondent is asked to apply the face cream and mentally note down the greasiness of the product.
- The respondents are again asked to wash and clean their face with a standard soap to remove all traces of the face cream applied. The face is then dried with a clean towel or tissue.
- The panellist is then given the second cream for application. The cream is applied in the similar fashion as in the first case.
- Each panellist is to note the greasiness of the face cream and report as to which of the two creams is greasier.

**Test Results obtained:**

| **Face cream found Greasier by the respondents** | **X** | **Y** | **Total "N"** |
|---|---|---|---|
| | **3** | **11** | **14** |

**Validity:**
The number of independent observations N equals 14. From the " Table[1] of probabilities associated with the values as small as observed values of *"X"* in a binomial test", we can see that the probability of obtaining a value of 3 for X with N=14 is 0.029, i.e., the probability of obtaining this result by chance is less than 3%.

**Conclusion:**

There is a strong evidence that sample 'Y" is perceived to be more greasy than sample "X"

**DIFFERENCE TEST**

---

[1] *see, the standard statistical table of probabilities in a binomial test*

This test is used to find out whether there is a difference between two products with respect to any specific attribute or characteristic. Difference test is extremely sensitive and can often indicate minute differences between two similar products. Various versions of difference test are used in the consumer product industry. The most common ones are the Triangle test, Duo-trio test, Dual standard test, Paired difference test, A not A test, and Multiple standard test. Triangle test is the most popular one and very widely used.

***1. TRIANGLE TEST***

**Example:**

Due to the rising cost of expeller grade of coconut oil a soap manufacturer want to change to Grade 2 solvent extracted coconut oil in the oil blend for making their soap base. Grade 2 solvent extracted coconut oil performs similarly like expeller grade oil. However, the soap base has a slightly stronger base odour. The manufacturer wants to know whether this base odour is masked by fragrance or is still detectable in the final soap made.

**Objective:**

To determine if soaps made by use of solvent extracted coconut oil can be detected by its base odour.

**Method to use:**

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

**Protocol and Standard operating procedure:**

- Thirty volunteers having an odour discriminating nose considered as good odour evaluators are selected as panellist.
- Two soap cakes made with expeller grade coconut oil (reference standard / control) are labelled "K2" and "M8" and one soap cake made out of solvent extracted grade 2 coconut oil (experimental) is marked "L7".
- The soap samples are all made having identical formulation, shape, size, and colour, except for the coconut oil type.
- The panellists are asked to smell the soap cake and select the odd one out of the three samples given.

**Test results obtained:**

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

**Validity:**

The function $\chi 2$ is given by

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

where O = observed value and E = expected value.

$\chi2 = \frac{(17-10)^2}{10} + \frac{(13-20)^2}{20}$

$\chi2 = 7^2 / 10 + (-7)^2 / 20$

$\chi2 = 49 / 10 + 49 / 20$

$\chi2 = (49 \text{ X } 2) + 49 / 20$

$\chi2 = (98 + 49) / 20$

$\chi2 = 147 / 20$

$\chi2 = 7.35$

Therefore, the calculated value of $\chi2$ is 7.35. This value is compared with the table[2] value of $\chi2$ for given degrees of freedom at a certain specified level of significance. (Generally 5% level is selected and used). If the calculated value of $\chi2$ is more than the table values of $\chi2$ the difference between theory and observation is considered to be significant, i.e., it could not have arisen due to fluctuations of simple sampling. If on the other hand the calculated value of $\chi2$ is less than the table value the difference between theory and observation is not considered as significant, i.e., it is regarded as due to fluctuations of simple sampling and hence ignored. The results though not statistically significant in the experimental observations may become statistically significant when carried out with larger number of sample observations. $\chi2$ are always positive with its upper limit as infinity. As $\chi2$ is derived from observations it is a statistics and not a parameter. The $\chi2$ test is also termed non-parametric where the test involve no assumption about the form of the original distribution from which the observation come. The test conducted has one degree of freedom, as there is only one constraint governing the data namely the total number of selection must be 30.

**Conclusion:**

From the $\chi2$ distribution table the value of $\chi2_{\ 0.05}$ with one degree of freedom i.e., v = 1 is 3.814. The calculated value of $\chi2$ is 7.35. Therefore the calculated value $\chi2_{\ 0.05}$ is higher than the table value. Therefore the difference between theory and experiment is significant and it could not have arisen due to fluctuations of simple sampling. The base odour in the soap made out of grade 2 solvent extracted coconut oil is detectable.

**As discussed earlier the alternative methods of sensory testing are as follows.**

## *2. DUO – TRIO TEST :*

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

The expected frequencies and so the contingency tables are as given below.

| | **Correct selection** | **Incorrect selection** | **Total** |
|---|---|---|---|
| **Experimental results** | Ox | Oy | N |
| **Expected results if no difference found (Hypothesis)** | N / 2 | N / 2 | N |

The function $\chi2$ is given by $\chi2 = \Sigma\, (O - E)^2 / E = (Ox - N / 2)^2 / N / 2 + (Oy - N / 2)^2 / N / 2$
where O = observed value and E = expected value.

## *3. DUAL – STANDARD TEST :*

In this case, four samples are used. Both one control and one experimental or different sample are identified to the subjects. Then the subjects are given two unknown samples and asked to pair them. The calculations are similar to the Duo-trio test.

---

[2] *see, the standard statistical tables of $\chi2$*

Now supposing the control and the experimental samples are not identified in the beginning, but presented together for pairing with two more samples that are similar the expected frequencies are altered to the standard triangle test. For example, If $P_1$ and $P_2$ are the first pair of samples and $Q_1$ and $Q_2$ the second pair and if there are no differences between them then the following pairs are possible.

**$P_1P_2 : Q_1\ Q_2$ ; $P_1Q_1 : P_2Q_2$ ; $P_1Q_2 : P_2Q_1$** of which **$P_1P_2 : Q_1\ Q_2$** is the only correct one.
The contingency table in this case can be given as below

| | Correct selection | Incorrect selection | Total |
|---|---|---|---|
| **Experimental results** | Ox | Oy | N |
| **Expected results if no difference found (Hypothesis)** | N / 3 | 2N / 3 | N |

The function $\chi 2$ is given by $\chi 2 = \Sigma\ (O - E)^2 / E = (Ox - N / 3)^2 / N / 3 + (Oy - 2N / 3)^2 / 2N / 3$
where O = observed value and E = expected value.

### *4. PAIRED DIFFERNCE TEST :*

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

### *5. "A" NOT "A" TEST :*

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

### *6. MULTIPLE STANDARD TEST :*

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

The contingency table in this case is as below.

| | Correct selection | Incorrect selection | Total |
|---|---|---|---|
| **Experimental results** | Ox | Oy | N |
| **Expected results if no difference found (Hypothesis)** | N / E | (E – 1) N / 3 | N |

Where E= number of control sample + 1 test sample.
The function $\chi 2$ is given by $\chi 2 = \Sigma\ (Ox - N / E)^2 / (N / E) + (Oy - (E - 1) N / E)^2 / (E - 1) (N / E)$

**MEDIAN TEST**

When there are considerable differences between control and experimental samples the median test is usually used for testing the variation between them. Median test is largely used where the attribute or property is not precise. i.e., psychometric.

**Example:**

Evaluation of gloss obtained on hair after use of a newly developed experimental shampoo in comparison with market product.
**Objective:**

To evaluate the newly developed shampoo “E” for gloss imparted on hair when compared with the market sample “M” of the rival company.

**Method to use:**

- An inexpensive laboratory, method is to be used as the product is in the early stages of development.
- The evaluation should be carried out on hair so that the gloss imparted can be observed.
- Salon testing is expensive so hair tresses may be used.

**Protocol and Standard operating procedure:**

- Take 12 matching tresses of hair from a common stock.
- The tresses are washed and cleaned before starting the evaluation.
- Each of the twelve hair tresses should be properly coded.
- Six tresses are used for experimental formulation and six for market shampoo.
- The tresses are used for shampoo treatment as per the normal method of usage.
- The hair tresses after using the shampoo is washed free of shampoo , dried and combed say 25 times using a standard comb.
- The tresses are kept on a tray each one separated from the other and presented to the respondents for evaluation.
- The evaluators are asked to rank the tresses according to the degree of gloss observed on the hair from after use of the shampoo.
- The average rank of the tresses is calculated and the results noted down for analysis.

| **Tress Number** | **AX** | **BW** | **CV** | **DU** | **ET** | **FS** | **GR** | **HQ** | **IP** | **JO** | **KN** | **LM** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Shampoo used** | **M** | **E** | **M** | **E** | **M** | **E** | **M** | **E** | **M** | **E** | **M** | **E** |
| **Respondent 1** | 3 | 4 | 7 | 5 | 8 | 1 | 10 | 6 | 12 | 9 | 11 | 2 |
| **Respondent 2** | 3 | 4 | 5 | 6 | 8 | 2 | 9 | 7 | 11 | 10 | 12 | 1 |
| **Respondent 3** | 2 | 4 | 3 | 5 | 9 | 8 | 10 | 7 | 11 | 6 | 12 | 1 |
| **Respondent 4** | 5 | 2 | 8 | 7 | 9 | 6 | 4 | 3 | 11 | 10 | 12 | 1 |
| **Respondent 5** | 3 | 4 | 8 | 5 | 9 | 2 | 7 | 6 | 12 | 10 | 11 | 1 |
| **Respondent 6** | 2 | 3 | 6 | 8 | 7 | 4 | 11 | 5 | 10 | 9 | 12 | 1 |
| **Respondent 7** | 2 | 4 | 8 | 3 | 6 | 5 | 11 | 7 | 10 | 9 | 12 | 1 |
| **Respondent 8** | 1 | 3 | 9 | 6 | 7 | 5 | 11 | 4 | 10 | 8 | 12 | 2 |
| **Respondent 9** | 1 | 2 | 9 | 7 | 8 | 5 | 6 | 4 | 11 | 10 | 12 | 3 |
| **Respondent 10** | 1 | 3 | 6 | 7 | 9 | 5 | 10 | 4 | 8 | 12 | 11 | 2 |
| | | | | | | | | | | | | |
| **Average Rank** | **2.3** | **3.3** | **6.9** | **5.9** | **8** | **4.3** | **8.9** | **5.3** | **10.6** | **9.3** | **11.7** | **1.5** |
| **RANK** | 2 | 3 | 7 | 6 | 8 | 4 | 9 | 5 | 11 | 10 | 12 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Shampoo E** | 1 | 3 | 4 | 5 | 6 | 10 |
| **Shampoo M** | 2 | 7 | 8 | 9 | 11 | 12 |

| | **High Gloss Observed** | | | | | | **MEDIAN** | **Low Gloss observed** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Rank** | **1** | **2** | **3** | **4** | **5** | **6** | | **7** | **8** | **9** | **10** | **11** | **12** |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Shampoo used** | **E** | **M** | **E** | **E** | **E** | **E** | | **M** | **M** | **M** | **E** | **M** | **M** |

**Validity:**

The probability of obtaining the above by chance can be calculated by use of the median test.

| Number | Experimental Shampoo | Market Shampoo | Total |
|---|---|---|---|
| **Better than median** | 5 ( p ) | 1 ( q ) | 6 |
| **Worse than median** | 1 ( r ) | 5 ( s ) | 6 |
| **Total** | 6 | 6 | 12 |

**Probability =** $$\frac{(p + q)!\ (r + s)!\ (p+r)!\ (q + s)!}{(p+q+r+s)!\ p!\ q!\ r!\ s!}$$

= 6 ! 6! 6 ! 6 ! / 12! 5! 1! 5! 1! (Note: 5! means 5 X 4 X 3 X 2 X 1 = 120)
= 3 / 77 = 0.0389 = 3.8 %
The probability of obtaining this rankings by chance is about 3.8 % i.e., approximately 4%

**Conclusion:**

There is strong evidence that the experimental shampoo gives more gloss on the hair after drying and combing when compared with the marketed product from competition.

**STUDENT ' t ' TEST:**

**1. To test the difference between the means of two samples**

The students 't" test is another useful classical statistical test used in consumer product evaluation. The technique is very useful when dealing with small samples and when the parent sample population is normal. Strictly speaking, therefore the results obtained by the "t" test will be true only for a normal population. Theoretical work however confirms that the results remain true for populations, which do not deviate markedly from normal distribution. However if we suspect that the population is markedly skewed that is 'U' or 'J' shaped then the 't' test cannot be applied with confidence. Given two independent set of samples, $n_1$ & $n_2$ with means $X_1$ & $X_2$ and standard deviations $S_1$ & $S_2$ then the significance of the difference between the two samples can be assessed by use of the student 't' test.

**Example:**

A new toilet soap fragrance formulation is developed to give more substantivity on skin when used. We already know from a previous study that our current soap available in the market is substantive on skin. We wish to claim that the newly developed product is better in terms of substantivity than the marketed product. It is not advisable to carry out expensive extensive salon consumer testing during the initial development stage. The relative substantivity of the new fragrance therefore have to be determined accurately only by laboratory testing .

**Objective:**

To find out whether the newly developed soap fragrance is better in term of substantivity in comparison to market sample.

**Method to use:**

- The company intends to replace the current soap in the market with soaps made with the new fragrance if tests prove it better in substantivity.
- A small but sensitive test is required before the company undertakes a larger market research study.
- Time is not a constraint, however the test should be reliable.
- Soaps made with new fragrance oil are to be tested with soaps procured from the market.

- The results can be validated by student 't' test.

**Protocol and Standard operating procedure:**

- Prepare two sets of soap samples one consisting of soaps made out of the newly developed fragrance and the other consisting of market sample used as control.
- Ideally, select ten to twelve volunteers (not less than six) who would use the soap samples sequentially during their bath and report on the fragrance substantivity on skin.
- The volunteers should be instructed to use the sample 'AY' for three days and self sniff out the time duration in minutes the fragrance lingers on their skin after bath. The sample 'KB' should be used on the subsequent three days and the fragrance lingering on skin noted down.
- The amount of rubs, soap used and lather generated on the body should be similar for both the soap samples issued to the respondents.
- The water temperature used for bathing should be same in both the cases.
- The average time duration in minutes, the fragrance lingers on the body for each sample and volunteer is noted down for analysis

**Results :**

| | **Substantivity on skin in minutes.** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Respondents** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **Sample AY** | 10 | 6 | 16 | 17 | 13 | 12 | 8 | 14 | 15 | 9 | * | * |
| **Sample KB** | 7 | 13 | 22 | 15 | 12 | 14 | 18 | 8 | 21 | 23 | 10 | 17 |

**Note:** *Two respondents due to personal reasons have failed to report on time, the results they have obtained on using sample AY.*

**Validity:**
Let us take the null hypothesis that the soap samples AY and XB do not differ significantly as regards its substantivity on skin.
Applying the 't' test

$$\text{'t'} = \frac{X_1 - X_2}{S}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$S = \sqrt{\frac{\Sigma(X_1 - X'_1)^2 + \Sigma(X_2 - X'_2)^2}{n_1 + n_2 - 2}}$$

where
$X_1$ = mean of the first sample. $X_2$ = mean of the second sample.
$n_1$ = number of observations of the first sample. $n_2$ = number of observations of the second sample.
S = combined standard deviation, and $(n_1 + n_2 - 2)$ is the degrees of freedom.

When the actual means are in fractions the deviations is taken from assumed means. In such a case, the combined standard deviation is obtained by the following formula.

$$S = \sqrt{\frac{\Sigma(X_1 - A_1)^2 + \Sigma(X_2 - A_2)^2 - n_1(X_1 - A_1)^2 - n_2(X_2 - A_2)^2}{(n_1 + n_2 - 2)}}$$

$X_1$ = Actual mean of the first sample. $X_2$ = Actual mean of the second sample.
$A_1$ = Assumed mean of the first sample. $A_2$ = Assumed mean of the second sample.
$n_1$ = number of observations of the first sample. $n_2$ = number of observations of the second sample.
S = combined standard deviation.

't' here is based on $(n_1 + n_2 - 2)$ degrees of freedom.

If the calculated value of 't' be greater than $t_{0.05}$ $(t_{0.01})$, the difference between the sample means is said to be significant at 5% (1%) level of significance, otherwise is said to be consistent with the hypothesis.

| Sample AY | | | Sample XB | | |
|---|---|---|---|---|---|
| **Substantivity On skin in minutes** | **Deviation from Mean 12 $(X_1 - X'_1)$** | $(X_1 - X'_1)^2$ | **Substantivity On skin in minutes** | **Deviation from Mean 15 $(X_2 - X'_2)$** | $(X_2 - X'_2)^2$ |
| **10** | -2 | 4 | **7** | -8 | 64 |
| **6** | -6 | 36 | **13** | -2 | 4 |
| **16** | 4 | 16 | **22** | 7 | 49 |
| **17** | 5 | 25 | **15** | 0 | 0 |
| **13** | 1 | 1 | **12** | -3 | 9 |
| **12** | 0 | 0 | **14** | -1 | 1 |
| **8** | -4 | 16 | **18** | 3 | 9 |
| **14** | 2 | 4 | **8** | -7 | 49 |
| **15** | 3 | 9 | **21** | 6 | 36 |
| **9** | -3 | 9 | **23** | 8 | 64 |
| * | * | * | **10** | -5 | 25 |
| * | * | * | **17** | 2 | 4 |
| $\Sigma$ **= 120** | $\Sigma(X_1 - X'_1) = 0$ | $\Sigma(X_1 - X'_1)^2 = 120$ | $\Sigma$ **= 180** | $\Sigma(X_2 - X'_2) = 0$ | $\Sigma(X_2 - X'_2)^2 = 314$ |

Mean amount of time duration the fragrance AY is substantive on skin $X_1 = \Sigma X_1 / n_1 = 120 / 10 = 12$ min.
Mean amount of time duration the fragrance XB is substantive on skin $X_2 = \Sigma X_2 / n_2 = 180 / 12 = 15$ min.

$$S = \sqrt{\frac{\Sigma(X_1 - X'_1)^2 + \Sigma(X_2 - X'_2)^2}{n_1 + n_2 - 2}}$$

$$S = \sqrt{\frac{120 + 314}{10+12-2}}$$

$S = 4.658$

Now $X_1 = 12$, $X_2 = 15$, $n_1 = 10$, $n_2 = 12$, $S = 4.658$, and $v = (n_1 + n_2 - 2) = (10 + 12 - 2) = 20$ therefore,

$$'t' = \frac{X_1 - X_2}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$'t' = \frac{12 - 15}{4.658} \sqrt{\frac{10 \text{ X } 12}{10 + 12}}$$

$'t' = 1.5$
at 5% level of significance the table[3] value of 't' = 2.09. The calculated value is 1.5, that is less than the table value. Therefore, the experiment provides no evidence against the hypothesis.

**Conclusion:**

We conclude that samples 'AY' and 'XB' do not differ significantly as regards their fragrance substantivity on skin.

**2. The difference test from zero.**

In case of paired data i.e., where we have the same sample but are trying to find out the effect of a certain attribute the 'difference test' is applicable. It is also possible to use a sample as its own control. In this case, if the treatment or handling has an effect there occurs a change in the value of the

[3] *see, the standard statistical table of 't'*

parameter that is measured. If there is no effect then the change in the value of the parameter measured will be zero. In this application of the 't' test we test whether the mean change in value is significantly different from zero or not. While applying this test, the value of 't' is obtained as given below.

$$'t' = \frac{d'-0\sqrt{n}}{S}$$

$$'t' = \frac{d'\sqrt{n}}{S}$$

Where d' = the mean of the difference
and S = the standard deviation of differences.

The value of 'S' is calculated as follows

$$S = \sqrt{\Sigma (d - d')^2 / (n - 1)} \text{ or } \sqrt{\Sigma d^2 - (d')^2 . n / (n-1)}$$

where "t' is based on (n-1) degrees of freedom.

**Example:**

The development chemist has developed a detergent powder formulation incorporating a Florescent-whitening agent (FWA) supposed to increase the brightness of the washed fabric. We have to confirm whether there is any effect of the FWA on the brightness of the washed cloth, before we go ahead in comparing the developed detergent with market samples that claim to give maximum brightness.

**Objective:**

To demonstrate that the FWA used in the detergent is capable of increasing the brightness of washed fabric.

**Method to use:**

- A sensitive but at the same time a relative simple test is required.
- Each sample is used as its own control in an in vitro test where conditions can be controlled precisely and easily. The variation from sample to sample in minimised and conditions maintained to achieved maximum sensitivity
- Washing by use of a tergo-to-meter and brightness measurement by reflectance.

**Protocol and Standard operating procedure:**

1. Take about 10 to 15 similar fabric samples of 10 X 10 centimetres.
2. Mark each sample swatches for identification with indelible marking ink.
3. Fold the swatches into four folds so that the background colour effect is minimised.
4. Take reflectance measurements of the swatches using a reflectance /emission spectrometer in accordance with the instrument suppliers instructions.
5. Note down the reflectance of the test fabrics.
6. Prepare 1% detergent solution at which concentration the washing test shall be carried out.
7. Measure out the required detergent solution by means of a measuring cylinder and pour into on of the tergo-to-meter beakers.
8. Similarly, fill five other beakers of the tergo-to-meter with standard detergent solution.
9. Switch on the water bath heater. Allow the temperature of the wash liquor to reach 30 degree centigrade and maintain this temperature throughout the test.
10. Introduce two swatches into each tergo-to-meter beaker and start the agitator.
11. Wash the cloth specimen for exactly 5 minutes at 30 degree centigrade, noting the time with a stopwatch.
12. Remove the beakers from the water bath and decant the solution. Squeeze the swatches lightly free of water. Rinse the beakers with fresh water and refit into the water bath.
13. Reintroduce the fabric swatches into the beaker and rinse for 1 minute by agitation in 500 ml of water. Decant the rinsed water and give a second rinse. Repeat the rinse a third time.
14. Dry the swatches at room temperature in the shade.

15. Measure the reflection of the washed swatches as before and note down the difference in readings obtained before and after washing.

**Results:**

| Swatches | Difference in Reflectance observed |
|---|---|
| 1 | 5 |
| 2 | 2 |
| 3 | 8 |
| 4 | –1 |
| 5 | 3 |
| 6 | 0 |
| 7 | –2 |
| 8 | 1 |
| 9 | 5 |
| 10 | 0 |
| 11 | 4 |
| 12 | 6 |

**Validity:**

Let us take a null hypothesis that the FWA used in the detergent does not increase the brightness of washed fabric.

Applying the difference test.

$$'t' = \frac{d' \sqrt{n}}{S}$$

$$S = \sqrt{\Sigma d^2 - (d')^2 . n / (n-1)}$$

| d | $d^2$ |
|---|---|
| 5 | 25 |
| 2 | 4 |
| 8 | 64 |
| -1 | 1 |
| 3 | 9 |
| 0 | 0 |
| -2 | 4 |
| 1 | 1 |
| 5 | 25 |
| 0 | 0 |
| 4 | 16 |
| 6 | 36 |
| **Σ d = 31** | Σ $d^2$ = 185 |

d'= Σ d / n = 31 /12 = 2.583

$$S = \sqrt{\Sigma d^2 - (d')^2 . n / (n-1)}$$

$$S = \sqrt{185 - (2.58)^2 . 12 / (11)} = \sqrt{(185 - 80.1) / 11} = \sqrt{9.54} = 3.09$$

Therefore d' = 2.583 , S = 3.09 and n = 12

$$'t' = \frac{2.583 \sqrt{12}}{3.09} = 2.9 \quad \text{where } v = (n - 1) = (12-1) = 11$$

for v =11 and at 5% level of significance the table value[4] of ‘t’ = 2.2

The calculated value of ‘t’ is 2.9 and is greater than the table value, which is 2.2. Hence, the results of the experiment do not support the hypothesis.

**Conclusions:**

From the experiment, we can conclude that the FWA used in the detergent in general increases the brightness / whiteness of the fabric washed. The result is significant at 5% level of significance, or there is 1 in 20 chance of getting this result accidentally with various experiments. The results are very encouraging for further testing with larger number of swatches and in real life situations with marketed detergents.

**ANALYSIS OF VARIANCE:**

One of the most powerful tools of statistical analysis is the analysis of variance. It consists of classifying and cross classifying statistical results and testing whether the means of a specified classification differ significantly. Analysis of variance enables us to analyse the total variation of our data into components, which may be attributed to various ‘sources’, or ‘causes’ of variation.

The ‘t’ test of the difference of means is an adequate procedure for testing the null hypothesis when we have means of only two samples to consider. In situations where we have three or more samples to be analysed at a time, then this can be achieved by the technique of analysis of variance. If the data is more extensive or of subsidiary divisions, then the analysis of variance as a technique is more convenient, unlike ‘t’ test which is convenient for quantitative measurements and when data is divisible into two groups.

In this technique, we assume that there is no difference between the test and control groups of samples and all results obtained are from samples that belong to the same homogeneous population. The population has a mean value, a normal distribution around the mean and a population variance that can be calculated. The analysis of variance test is not meant to test for the significance of the difference between two sample variations, but rather its purpose is to test for the significance of the differences among sample means. This is done by use of the mechanism of the ‘F’ – test, for testing for the significance of the difference between variances. The test is designed so that the variances being compared are different only if the means under consideration are not homogeneous. Significant values of ‘F’ indicate that the means are significantly different from one another.

The value of F is computed with the variance between samples means as the numerator and the variance within the sample means as the denominator. The calculated value of F is compared with the table value of F for the required degrees of freedom at a critical level (generally 5%) level of significance. If the calculated value of F is greater than the table value, it is concluded that the difference in sample means is significant. This means that it could not have arisen due to fluctuations of simple sampling or the sample do not belong to the same population. On the other hand if the calculated value of F is less than the table value the difference is said to be not significant and due to fluctuations of simple sampling.

The total population variance can be actually broken down into different components for analysis. The variability between groups is significantly larger than the variability between them. This variability is also called error. If however it is the other way, that is the variability between samples (error) is larger than the variability between groups then we can consider that there is a significant difference between groups and they do not all belong to the same population.

The technique of analysis of variance can be carried out separately for **one-way classification** and **two-way classification**. The various steps to be followed are as described below.

- **ONE-WAY CLASSIFICATION**

**Example:**

[4] *See, the standard statistical table value of ‘t’*

The formulation development lab has developed a deodorant that is claimed to keep body malodour away for a longer time. It has to be tested against the two major competitions available in the market along with the companies current deodorant sold in the market. The newly developed deodorant is hoped to be significantly better than the company's market brand and is expected to replace it for fuelling growth and market share in the deodorant segment.

**Objective:**

To evaluate whether the newly developed deodorant 'E4' is better than the competition brands 'K3' and 'L6' and not significantly worse than the companies current marketed product 'C2' when used in the conventional method by consumers.

**Method to use:**

1. The test must be carried out in vivo so that it stimulates actual usage practices.
2. A four-cell test is required; Newly developed product 'E4', current product 'C2', and two market products ' K3' and 'L6'.
3. We already know that the current product 'C2' performs equally well if not better than the competition market product.
4. The present test therefore should be sensitive to confirm this along with the ability to evaluate whether the newly developed deodorant is better or not.

**Protocol and Standard operating procedure:**

1. Select a panel of 20 volunteers.
2. The volunteers are instructed not to use any deodorant or antiperspirant product under their arms except the ones provided for the test.
3. Non medicated soap samples are supplied to all volunteers so that they can maintain the minimum required standard of cleanliness.
4. The volunteers are instructed not to go swimming or use products that might interfere with the test. E.g., depilatories.
5. Two weeks wash off phase is provided to each volunteer so that effect of any active products previously used is eliminated.
6. All the volunteers are first provided placebo products for one week so that the base line odour scores are generated.

***The procedure to be followed can be briefly described as below.***

- The volunteers report at the evaluation centre at an acceptable fixed time every day.
- A provision for a separate room where in the subject can undress in seclusion to provide access to their axillae is made.

- The odour under the left arms is rated using the following 9-point scale.

Very 1 | 2 3 4 5 6 7 8 | 9
Very
Weak Weak Moderate Strong
Strong

- The placebo / experimental product is applied to the right axilla and rated using the same 9-point scale as given above.

- The odour assessment is done by direct sniff test by self to avoid embarrassment to both subject and the evaluator.

- The time duration upto which body malodour is absent both qualitative and quantitative is noted down.

- The volunteers dress again and check out of the evaluation centre, after being asked to report back after a fixed time interval for reassessment.

- In case of base line odour generation, reassessment is done after 5 hours for convenience.

- The deodorant is not reapplied during any subsequent assessment for the same day.

- The scores are averaged for each axilla over a one-week period to provide base line scores for individual volunteers.

7. Based on the base line scores the volunteers are distributed into four groups, of 5 volunteers, so that each group has the same or similar overall mean score and similar distribution of scores.
8. Each axilla is regarded as an individual test site and assigned to groups.
9. Each subject is provided with one product for use under left arm and a different one on the right arm.
10. The odour assessment is carried out as described and repeated for a minimum one week with each of the deodorants. The scores are averaged for each axilla over the one-week period to provide the scores for individual volunteers.

**Results:**

| **Maximum time duration in hours during which there is an absence of body malodour in the test sites by using the deodorant samples** | | | |
|---|---|---|---|
| **E4** | **C2** | **K3** | **L6** |
| 4 | 6 | 9 | 6 |
| 5 | 5 | 6 | 4 |
| 6 | 4 | 8 | 6 |
| 4 | 7 | 3 | 8 |
| 3 | 2 | 4 | 7 |
| 4 | 6 | 9 | 7 |
| 5 | 6 | 6 | 5 |
| 6 | 5 | 8 | 6 |
| 4 | 7 | 3 | 8 |
| 4 | 2 | 4 | 8 |

**Validity:**

Let us take a hypothesis that the four deodorants used in the tests do not differ significantly during normal consumer use.

We apply the analysis of variance test as follows.

| **E4** | | **C2** | | **K3** | | **L6** | |
|---|---|---|---|---|---|---|---|
| $X_1$ | $(X_1)^2$ | $X_2$ | $(X_2)^2$ | $X_3$ | $(X_3)^2$ | $X_4$ | $(X_4)^2$ |
| 4 | 16 | 6 | 36 | 9 | 81 | 6 | 36 |
| 5 | 25 | 5 | 25 | 6 | 36 | 4 | 16 |
| 6 | 36 | 4 | 16 | 8 | 64 | 6 | 36 |
| 4 | 16 | 7 | 49 | 3 | 9 | 8 | 64 |
| 3 | 9 | 2 | 4 | 4 | 16 | 7 | 49 |
| 4 | 16 | 6 | 36 | 9 | 81 | 7 | 49 |
| 5 | 25 | 6 | 36 | 6 | 36 | 5 | 25 |
| 6 | 36 | 5 | 25 | 8 | 64 | 6 | 36 |
| 4 | 16 | 7 | 49 | 3 | 9 | 8 | 64 |
| 4 | 16 | 2 | 4 | 4 | 16 | 8 | 64 |
| $\Sigma = 45$ | $\Sigma = 211$ | $\Sigma = 50$ | $\Sigma = 280$ | $\Sigma = 60$ | $\Sigma = 412$ | $\Sigma = 65$ | $\Sigma = 439$ |

The sum of all the items of various samples = $X_1 + X_2 + X_3 + X_4$, = 45 + 50 + 60 + 65 = 220

Correction factor = $T^2/N = (220)^2 / 40 = 48400 / 40 = 1210$.

The total sum of the squares = $(X_1)^2 + (X_2)^2 + (X_3)^2 + (X_4)^2 - T^2/N$

= 211 + 280 + 412 + 439 – 1210 = 1342 – 1210 = 132.
Sum of the squares between the samples is obtained as follows

$= (X_1)^2 / N + (X_2)^2 / N + (X_3)^2 / N + (X_4)^2 / N - T^2/N$

$= (45)^2 / 10 + (50)^2 / 10 + (60)^2 / 10 + (65)^2 / 10 – 1210$

= 12350 /10 –1210 = 25

Sum of the squares within samples = 132 – 25 = 107
The above results can be tabulated as follows.

| Source of variation | Sum of squares | Degrees of Freedom | Mean square |
|---|---|---|---|
| **Between Samples** | 25 | 3 | 8.33 |
| **Within Samples** | 107 | 36 | 2.97 |
| **Total** | 132 | 39 | |

F = Variance between samples / Variance within samples = 8.33 / 2.97 = 2.80

From the table[5] of significance points of variance at 'F' for $v_1 = 3$ and $v_2 = 36$ at 5% level of significance the critical value of F is 2.80. This is less than the table value and hence the difference in the mean values of the sample is not significant. The samples could have come from the same universe and the difference is only due to fluctuations of simple sampling.

**Conclusion:**

The newly developed deodorant is not significantly better than either our current product or that of the competition when used in a conventional manner by consumers.

- **TWO-WAY CLASSIFICATION**

In a two-way classification, the data are classified according to two different factors. The procedure for applying the analysis of variance is different from the one way classification and is of the following form.

| Source of variation | Sum of squares | Degrees of Freedom | Mean square |
|---|---|---|---|
| **Between Columns** | SSC | (c-1) | MSC = SSC / (c-1) |
| **Between Rows** | SSR | (r-1) | MSR = SSR / (r-1) |
| **Residual** | SSE | (c-1)(r-1) | MSE = SSE / (c-1) (r-1) |
| **Total** | SST | N-1 | |

SSC = Sum of the squares between columns.
SSR = Sum of the squares between rows.
SSE = Sum of the squares for the residual.
SST = Total sum of squares.

- The sum of the squares for the source 'residual' is obtained by subtracting from the total sum of squares, the sum of squares between column and rows.
- The total number of degrees of freedom = (cr-1) where 'c' refers to the columns and 'r' refers to the rows.

[5] *see, the standard statistical table value of 'F'*

- The total sum of the squares, sum of squares for 'between columns' and sum of squares for 'between rows' are obtained in the same manner as before.
- Residual (Error or remainder) = Total sum of squares – square for between columns – square for between rows.

In the problem involving two-way classification 'residual' is the measurement for testing the significance, it represents the magnitude of variations due to the forces called 'chance'.

**Example:**

There are two opinions about the strength of the fragrance perceived in soaps made in the Soap pilot plant and in the factory, with and without vacuum plodding. We have to find out whether there is any marked difference perceived in fragrance impact due to the method of manufacture.

**Objective:**

To find whether the method of manufacture of soap has an effect on the strength of the fragrance as perceived by consumers.

**Method to use:**

Soaps are manufactured by use of the various methods of manufacture e.g.,

1. In the pilot plant
2. In a plodder without vacuum
3. Using a vacuum plodder.

The finished soaps after maturity is evaluated with respondents who are asked to rate the soap made on the basis of the strength of the fragrance perceived by them and results analysed.

The soap batches are replicated at different locations in the same manner to confirm and validate the results.

**Results:**

| ⇓ Method of manufacture | Average scores obtained | | | | |
|---|---|---|---|---|---|
| **Batches ⇒** | **1** | **2** | **3** | **4** | **5** |
| **Plodding without vacuum** | 26 | 27 | 30 | 21 | 26 |
| **Plodding with vacuum** | 23 | 26 | 22 | 20 | 33 |
| **Pilot Plant** | 33 | 36 | 29 | 36 | 42 |

**Validity:**

Let us take the hypothesis that there is no difference between the soap samples with regards to the mode of manufacture. For the sake of simplifying the calculations, subtract 25 from the given data. Coding the above data, we get the following.

| ⇓ Method of manufacture | Average scores obtained | | | | | Total |
|---|---|---|---|---|---|---|
| **Batches ⇒** | **1** | **2** | **3** | **4** | **5** | |
| **Plodding without vacuum** | 1 | 2 | 5 | -4 | 0 | **4** |
| **Plodding with vacuum** | -2 | 1 | -3 | -5 | 8 | **-1** |
| **Pilot Plant** | 8 | 11 | 4 | 11 | 17 | **51** |
| **Total** | **7** | **14** | **6** | **2** | **25** | **54** |

Here Total = 54 and N = 15.

Correction factor = $T^2/N = (54)^2/15 = 2916/15 = 194.4$

- The total sum of the squares = $\dfrac{(4)^2 + (-1)^2 + (51)^2}{5} - T^2/N$

$= (16 + 1 + 2601)/5 - 194.4 = 523.6 - 194.4 = 329.2$ where degrees of freedom, $v = 3-1 = 2$

- Sum of the squares between soap fragrance batches

$= (7)^2/3 + (14)^2/3 + (6)^2/3 + (2)^2/3 + (25)^2/3 - T^2/N$

$= (49)/3 + (196)/3 + (36)/3 + (4)/3 + (625)/3 - 194.4$

$= 910/3 - 194.4 = 108.9$ here degrees of freedom is $(r-1) = (5-1) = 4$

- Total sum of squares.

$= (1)^2 + (-2)^2 + (8)^2 + (2)^2 + (1)^2 + (11)^2 + (5)^2 + (-3)^2 + (4)^2 + (-4)^2 + (-5)^2 + (11)^2 + (0)^2 + (8)^2 + (17)^2 - T^2/N$

$= (1 + 4 + 64 + 4 + 1 + 121 + 25 + 9 + 16 + 16 + 25 + 121 + 0 + 64 + 289) - T^2/N$

$= 760 - 194.4 = 565.6$

- Residual or Remainder = Total Sum of Squares – Sum of squares between method of manufacture – sum of squares between fragrance batches

$= 565.6 - 329.2 - 108.9 = 127.5$

here the degrees of freedom for remainder is $14 - (5-1) - (3-1) = 8$

Analysis of variance Table is of the following form.

| Source of variation | Sum of squares | Degrees of Freedom | Mean square | F |
|---|---|---|---|---|
| **Between Batches** | 108.9 | 4 | 27.225 | $F_1$ : 27.225 / 15.937 = 1.708 |
| **Between methods of manufacture** | 329.2 | 2 | 164.6 | $F_2$ : 164.6 / 15.937 = 10.33 |
| **Residual** | 127.5 | 8 | 15.937 | |
| **Total** | 565.6 | 14 | | |

From the table[6] of significance points of variance for degrees of freedom 4 and 8, $F_{0.05} = 3.84$ and for degrees of freedom 2 and 8, $F_{0.05} = 4.46$ at 5% level of significance

**Conclusion:**

1. The calculated value of $F_1$ is 1.708. This is less than the table value of 3.84. We conclude that there is no significant difference between the various batches of soap made.
2. The calculated value of $F_2$ is 10.33. This is greater than the table value of 4.46. Therefore, we conclude that the soaps are significantly different with regard to the mode of manufacture.

[6] *see the standard statistical table value of 'F'*

**EXPERIMENTAL DESIGN:**

We have seen how to analyse and interpret the experimental test results when obtained. This analysis and interpretation will depend mostly on the method or way the results have been obtained. The observation obtained after a long and laborious experimentation may not yield any useful information. To avoid such eventualities the question of design of experiment and the method of analysis of observation should be decided and confirmed at the same time.

The experimental design that is important as the analysis has to consider the following factors.

1. ***The range of relative induction:***

The design of any experiment depends on the type of problem and the type of inference we have to make. In order to get a general inference the sample observation should be taken from a hypothetical infinite population. The method of sampling should be random and not biased or to suit the convenience. For example, if a product is created for the rural market then the test experiment should be carried out more on the rural consumers and not on respondents based in cities to make inferences more reliable and authentic.

2. ***Null hypothesis:***

When inferences are drawn on some common features present in the observed facts then the inferences drawn is called as ‘hypothesis’. When setting up the experimental design one control and one experimental groups are always set up. Namely one group without the factor say ‘x’ (Control) and an other group with the factor ‘x’ (Experimental). Statistical analysis of test are always concerned with the difference between the observations (control and experimental) and not with the isolated observation. In any experiment carried out control observation is equally important along with the experimental observation. It is important that when the difference in observation is considered all other factors except the factor, ‘x’ is eliminated, so that the results obtained is due to the factor ‘x’ only. Sometimes there is one hypothesis that two or more sets of measurements are likely to have been drawn by chance from the same population. This can be tested with statistical methods. Therefore, according to this hypothesis the observed difference is likely to be due to errors of sampling as there is no conformable difference between the groups. This hypothesis is called as null hypothesis and is especially suitable for rigid deduction and practical use.

3. ***Segregation of causes of the variations:***

When experimental design is considered we should make sure that the results or differences obtained in the group containing factor ‘x’ is due to the presence of the factor ‘x’ only. However it is possible that a larger number of other minor factors e.g., age, sex, physical condition of respondents will effect the experiment. We are also not sure that these small or minor differences will not effect the observed results. Such small variations in between the subjects should be avoided as far as possible. The experiment should be so done that non-specific factors that may affect the result contributing to the residual variance do not affect the variance related to the factor ‘x’. This is usually achieved by random allocation.

4. ***Random allocation*:**

The selection of the experimental subjects should be made at random. Not all subjects are the same. They all have their strengths and weakness in different areas. Subjects having strengths in one area and weakness in another area and vice versa should not be placed in the same group. The groups selected for the experiments should be equally distributed and contain subjects belonging to all categories. The groups are equal and not lopsided with one group of respondents stronger than the others are.

**REDUCTION OF EXPERIMENTAL ERROR:**

In any experiment designed, the significance of the result will depend on the magnitude difference produced by the factor ‘x’. That is on the difference between the experimental results and control results, on the number of observation and inversely on the error or standard deviation. For a reliable inference, the number of observations should be as high as practically possible. The experimental material should be homogenous and conditions rigidly controlled. Statistical analysis should be preferred to estimate the cause of variation, not attributable to experimental factors.

The experiments should be conducted repeatedly in identical conditions to reduce any experimental error and to give an inference that is precise and accurate. Factors like age, sex, temperature, etc., that may affect the experimental observations should not be ignored. The experimental and control tests should be held in similar conditions by proper selection of subjects to increase the validity of the data.

**CONCLUSION:**

We know two primary approaches; namely, experimental methods and statistical methods are employed for advancing our knowledge. Experimental methods are without doubting more scientific and historically more often used. The cause and effect relations are often established and investigated in a controlled set-up. However, due to various constraints in some cases, experimental methods cannot be applied. We resort to statistical methods in such cases. In applying statistical methods, the problem is studied as systematically as in the experimental methods, although the system is not the same. The result though not as accurate as the experimental method is decidedly better than having no results. It should however be noted that the distribution between the experimental method and the statistical methods is formal and arbitrary and should not be taken as rigid and definite. In practice scientists often, combine elements of both methods to solve the problem. Which technique should be applied in a particular situation will depend on the problem or the object of investigation. The researchers can utilise various techniques, each of which is appropriate to a specific type of situation.

Objective tests do not experience subjective bias. This helps the development chemists to formulate a product with specific properties and set up protocol for tests that are based on scientific parameters and knowledge. Consumer likes and their needs are most important to be fulfilled if we have to be successful in the consumer product industry. This aspect should be the basis on which objective tests are designed. The results obtained and information assimilated should help the consumers subjectively appraise the product themselves and lead to their ultimate satisfaction.

**NOTE:**

This paper gives an over view of consumer product evaluation and briefly explains the theory behind the various methods used in product evaluation. It also illustrates the method to use the principles enumerated making use of identifiable live examples, also ensuring that the selections made are precise and accurate. Illustrations are selected to help understand the techniques described so that the tools learnt can be successfully applied by laboratory personnel in their routine roles. The views expressed in this paper are solely those of the author and not necessarily of the organisation he is employed or that of the publisher. There exists many other ways to carry out the evaluation of products and the method given in this paper is only one of the ways. The paper intended to a non-statistical non-expert does not attempt to cover all aspects of statistical techniques used in consumer product evaluation. Readers interested in knowing more details about the statistical techniques are advised to read standard works on statistical methods of analysis available in the market.