i'Tec_H3
August 2001

Keywords: Mix/Mixing , Homogeneity , Statistics , General

$CV_{homogeneity}$ value calculation Application of the random-effects model for variance analysis

Total variance determined on the basis of variations in concentrations between different samples provides a great deal of information, which is not all connected with the mixer's or production line's performance. The most important information appears to be that generated by the entire analytical process: sampling technique, transport, laboratory processing of samples, size of test specimens, dilution, extraction, etc.
In line with the approach provided by Cline (1978), Tecaliman has chosen to systematically test intra-sample variability in comparison with inter-sample variability.
In order to properly apply this approach, it must be accepted that there will always be a time when the variation in concentration generated by the mixing process will be lower than that determined by analysis. Therefore, in other words, there is always a moment when it is impossible to test the performance of a mixer (Yeung and Hersey, 1979 - Heidenreich, 1998). Consequently, the analytical process using a homogeneity tracer must be as precise as possible and it is advisable to measure whether this objective has been achieved.
As a result, the approach requires a comparison of the variance explained by the tested factor (the samples) with the variance originating from a number of analyses conducted on each of the samples to be taken as the basis, in order to provide a picture of the effect of the analytical process.
The statistical technique used to do this is variance analysis. There are two models, one known as "fixed" and the other known as "random", but this term does not refer to the random distribution of a product. The random model was chosen, following discussions with statisticians, as it is more closely in line with the process examined.

1. Theoretical model

This model is described by Snedecor and Cochran (1971). In our field, whether this model is fixed or random, variance analysis takes the same sample and analysis pattern as its basis. A homogeneity control test consists of taking n samples from a powder mix and of taking **p** tracer concentration measurements for each of the samples.
Each concentration measurement is identified by the value X_{ki}, **k** is a variant from 1 to **n**, and **i** is a variant from 1 to **p**. In the case of tests carried out by Tecaliman, p = 2 where i varies between 1 and 2. Therefore, the sampling and analysis pattern is that in Figure 1.

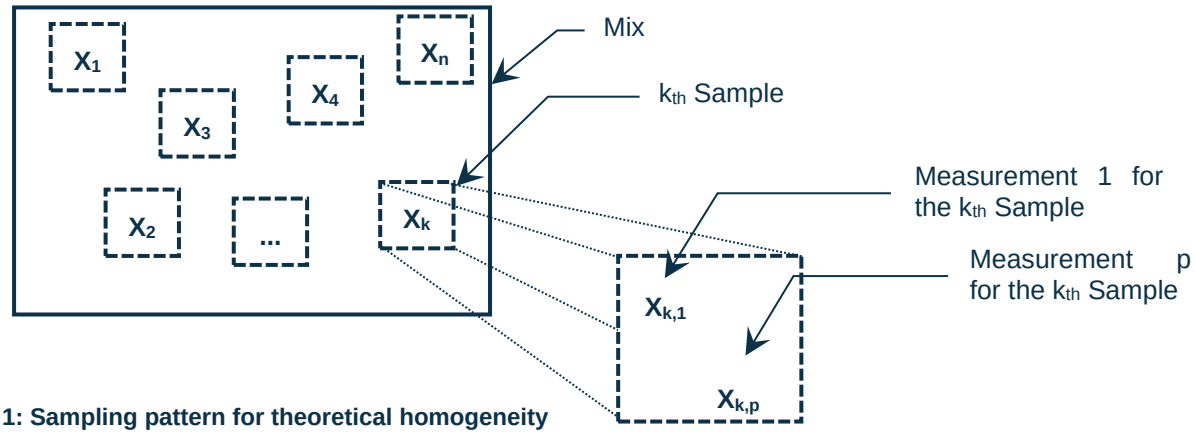

Figure 1: Sampling pattern for theoretical homogeneity control

Within this framework, **pn** measurements are obtained. The random model reflects the fact that all the samples collected from a mix are only a set of samples among others that it would have been possible to collect. Conducting a study using a "fixed" model would assume that the set of samples collected provides a perfect picture of all the possible sets and of the entire population itself. The random model does not make this assumption and attempts to take account of errors made when choosing samples.
The following equations enable the calculations performed to be precisely described. In this model, the change in concentration measurements X_{ki} is provided by:

$$X_{ki} = \mu + A_k + \varepsilon_{ki}$$

or

$$A_k = N\ (0, \sigma_A),\ \varepsilon_{ki} = N\ (0, \sigma)$$

- μ = mean
- A_k is the difference between the concentration of the sample k and the mean concentration for the population μ. This term reflects the existence of a variation in concentration between each of the samples. Each sample has its own A_k value. This is a 0 mean random variable as it is defined as a deviation from the mean μ. For the sake of simplicity, it is assumed that the A_k values have a normal mean distribution O and standard deviation σ_A, scored $N\ (0, \sigma_A)$.
- ε_{ki} reflects:
  - measurement errors
  - the variation in concentration between test specimens from the same sample.

The A_k values and ε_{ki} values are presumed to be independent. It is assumed that the ε_{ki} values have a normal zero mean distribution and standard deviation σ scored = $N\ (0, \sigma)$.
In the case of a mix, the aim will be to estimate σ_A , which, in this case, corresponds to the standard deviation for concentration of the product between samples, leading to a correct estimate of the coefficient of variation originating from the homogenisation operation.

Variance analysis results in the following parameters being determined:

- Total mean μ

$$\mu = \frac{\sum_{k=1}^{n}\sum_{i=1}^{2} X_{ki}}{pn}$$

- The sum of squares of variations between samples, SCE:

$$SCE = p\sum_{k=1}^{n}\left(\overline{X}_{ki} - \mu\right)^2$$

  SCE has p-1 degrees of freedom, i.e. one degree in the case of two analyses per sample.

- The sum of squares of intra-sample variations, SCI:

$$SCI = \sum_{k=1}^{n}\sum_{i=1}^{p}\left(X_{ki} - \overline{X}_{k.}\right)^2$$

  SCI has 2n-n degrees of freedom, therefore n degrees.

- Total variance V_T:

$$V_T = \frac{SCE}{p(n-1)} + \frac{SCI}{pn}$$

- The mean square between samples or inter-sample variation is estimated by CME:

$$CME = \frac{SCE}{n-1}$$

- The mean intra-sample square or intra-sample variation is estimated by CMI

$$CMI = \frac{SCI}{n(p-1)}$$

A comparison between the inter-sample variation (CME) and the intra-sample variation (CMI) is performed using a Fisher's exact test. This comparison is made using an $F_{calculated}$ value that can be compared with a Fisher-Snedecor table (Snedecor and Cochran, 1971) depending on the risk in question:

$$F_{calculé} = \frac{CME}{CMI}$$

Table 1 provides some $F_{theoretical}$ values based on the number of samples for two analyses conducted per sample.
If the $F_{calculated}$ value is higher than the $F_{theoretical}$ value, it is possible to state that there is a significant difference between the two sources of variation. If the $F_{calculated}$ value is lower than or equal to the $F_{theoretical}$ value, the two sources of variation do not differ significantly.

No. of samples	$F_{theoretical}$
5	5.19
10	3.02
15	2.42
20	2.13
25	1.96
30	1.85

Table 1

The existence of a significant difference makes the resulting CV calculation more robust, as the analysis method used, in this case, is sufficiently accurate to observe the differences between samples. In the end, three variances can be calculated: total variance, residual variance and homogeneity variance. These variances enable the corresponding coefficients of variation to be calculated.
Total variance remains the same. However, according to the random model, the mean intra-sample square is an estimate of the variance in ε_{ki} values:

$$\mathbf{CMI} = \sigma^2$$

and the mean square between samples (CME) is not a direct estimate of the variance in A_k values, but an estimate of the sum:

$$\mathbf{CME} = \sigma^2 + \mathbf{p}\sigma_A^2$$

Consequently, an estimate of the variance in A_k values can be obtained by:

$$\sigma_A^2 = \frac{CME - CMI}{p} = V_T - CMI$$

The coefficient of variation between samples, known as “homogeneity” is obtained by:

$$\mathbf{CV_{hom.}} = \mathbf{100}.\frac{\sqrt{\sigma_A^2}}{\mu}$$

Application of the random model may result in a residual variance that is higher than the total variance, which has the consequence of resulting in a negative homogeneity variance (obtained from the difference in the two others). In this case, the variance and the related coefficient of variation are set as equal to 0 by convention.
Consequently, each study of the homogeneity of a feed (for which at least two analyses are conducted per sample) results in four parameters being determined:

- The total mean.
- The significance of the variance analysis.
- The $CV_{Homogeneity}$ value (Standard deviation of the sample factor compared with the total mean).
- The CV_{total} value (Overall standard deviation compared with the total mean).

The coefficient of variation has the advantage of being a dimensionless number that, in theory, can enable results to be compared with each other. Nevertheless, there are drawbacks that should be remembered:

- **It depends on the mean**: if the latter decreases, with consistent standard deviation, the CV increases. Therefore, this parameter must always be verified, before continuing with the study. Consequently, carry-over has a potential impact on homogeneity and vice versa.
- **It depends on standard deviation**: validly determining this value requires a population of samples that is larger than that needed to determine a reliable mean.
- **It cannot be totalled.** Unlike variance, it is not possible to add coefficients of variation together. This sole difference leads many statisticians to prefer variance to coefficients of variation as an indicator of heterogeneity.

2. Examples

Take a mix into which a tracer has been incorporated at a level of 100 ppm. 10 samples are taken from the mix and the tracer concentration is analysed using two test specimens from each of the samples. As a result, two concentrations are obtained per sample (Table 2).

	Analyses	
Samples	**1**	**2**
1	99	98
2	95	97
3	102	98
4	101	102
5	97	100
6	89	92
7	103	101
8	101	99
9	94	96
10	92	90

Table 2: Example of results

The overall mean **μ** is **97.3 ppm**. Therefore, it is lower than expected but will be used to calculate the CV. We also obtain:

- SCE = **300.2**
- SCI = **28.0**

- V_T = 18.078
- CME = 300.2/9 = **33.356**
- CMI = 28.0/10 = **2.800**

Consequently, the $F_{calculated}$ value = 33.356/2.8 = **11.9** is higher than the $F_{theoretical}$ value of **3.02** (obtained for a risk of 5 % and degrees of freedom of 9 for the numerator and 10 for the denominator). Therefore, the variation between samples is significantly higher than the intra-sample variation.

Consequently, the variance in A_k values is:

$$\sigma_A^2 = \frac{\mathbf{CME - CMI}}{\mathbf{p}} = \frac{33.356 - 2.8}{2} = 15.278$$

and the homogeneity coefficient of variation is:

$$CV_{hom.} = 100.\frac{\sqrt{\sigma_A^2}}{\mu} = 100.\frac{\sqrt{15.278}}{97.3} = 4.02\ \%$$

Calculating the total variance enables a coefficient of variation of 4.4 % to be obtained and calculating residual variance (CMI) produces a coefficient of 1.7 %. These results clearly demonstrate the non-additivity of coefficients of variation. This is a very simple example for which the analysis used is high quality.

In the following example (Table 3), the overall mean is still 97.3 ppm.

	Analyses	
Samples	**1**	**2**
1	106	94
2	99	90
3	109	94
4	105	95
5	104	96
6	93	85
7	110	97
8	105	92
9	101	92
10	96	83

Table 3: Example of results

Nevertheless, the calculation parameters become:

- SCE = **435.2**
- SCI = **633.0**
- V_T = 55.828
- CME = 435.2/9 = **48.36**
- CMI = 633.0/10 = **63.3**

In this case, the $F_{calculated}$ value becomes 0.76 and is therefore lower than the $F_{theoretical}$ value of 3.02. The intra and inter sample variation are not significantly different. A simple comparison of the CME and CMI clearly shows that they are of the same order.
It is possible to conclude that the homogeneity coefficient of variation is invalid when compared to that generated by the analytical procedure.
If calculation is continued, the variance in A_k values obtained is negative, which is characteristic of this model:

$$\sigma_A^2 = \frac{\mathbf{CME - CMI}}{\mathbf{p}} = \frac{48.36 - 63.3}{2} = -7.47$$

In this case, total variance can be calculated (7.7 %) but it should be noted that it is entirely an expression of variations generated by the analytical procedure.

3. Conclusion

The advantage of applying this model is that it is uncompromising as regards the quality of the analytical procedure. It rules out a misinterpretation of the actual effect, the homogenisation operation or that of a stimulating factor, which could result in variations in homogeneity (physical characteristic of the same additive, mixing time, etc.). It is also more rigorous from a statistical point of view.

4. Bibliography

Cline A.L., 1978. Statistical evaluation of feed uniformity. Feedstuffs, 26TH JUNE, 40-41.

Yeung C.C., Hersey J.A., 1979. Ordered powder mixing of coarse and fine particulate systems. Powder Technol., 22, 127-131.

Heidenreich E., 1998. Microingredients in the mixer. Feed Tech., 2, 2, 13-17.

Snedecor G.W., Cochran W.G., 1971. Méthodes statistiques. ACTA. Sixth edition.