Mots-clés : Mélange, Homogénéité, Statistique, Généralité

Calcul du cv$_{\text{homogénéité}}$
Application du modèle aléatoire de l'analyse de variance

Une variance totale déterminée à partir des variations de concentrations entre différents échantillons est porteuse de multiples informations qui ne sont pas toutes liées aux performances de la mélangeuse ou de la ligne de fabrication. Les plus importantes semblent celles engendrées par l'ensemble du processus analytique : technique de prélèvement, transport, traitement des échantillons au laboratoire, taille de prises d'essais, dilution, extraction, ...
Dans la logique apportée par Cline (1978), Tecaliman a fait le choix de tester systématiquement la variation intra-échantillon par rapport à la variation inter-échantillon.
Pour bien aborder cette démarche, il faut admettre qu'il existera toujours un moment où la variation de concentration engendrée par le processus de mélange sera plus faible que celle perçue par l'analyse. En d'autres termes, il existe donc toujours un moment où il est impossible de tester les performances d'une mélangeuse (Yeung et Hersey, 1979 - Heidenreich, 1998). Le processus analytique du traceur d'homogénéité doit donc être le plus précis possible et il convient de mesurer si cet objectif a été atteint.
La démarche nécessite donc de prendre pour base la comparaison de la variance expliquée par le facteur testé (les échantillons) à la variance issue de plusieurs analyses pratiquées sur chacun des échantillons, image de l'effet du processus analytique.
Pour ce faire, la technique statistique utilisée est celle de l'analyse de variances. Il en existe deux modèles l'un dit « fixe » , l'autre dit « aléatoire », mais ce terme ne fait pas référence à la distribution aléatoire d'un produit. Le modèle aléatoire a été choisi, suite à des entretiens avec des statisticiens, car il est plus rigoureusement en accord avec le procédé étudié.

1. Modèle théorique

Ce modèle est décrit par Snedecor et Cochran (1971). Dans notre domaine, que ce modèle soit fixe ou aléatoire, l'analyse de variance prend pour base le même schéma de prélèvement et d'analyse. Un test de contrôle d'homogénéité consiste à prendre **n** échantillons dans un mélange de poudre et à faire **p** mesures de concentration d'un traceur dans chacun des échantillons.
Chaque mesure de concentration est identifiée par la valeur X_{ki}, **k** variant de 1 à **n**, et **i** variant de 1 à **p**. Dans le cas des essais réalisés par Tecaliman, p = 2 dont i varie de 1 à 2. Le schéma de prélèvement et d'analyse est donc celui de la Figure 1.

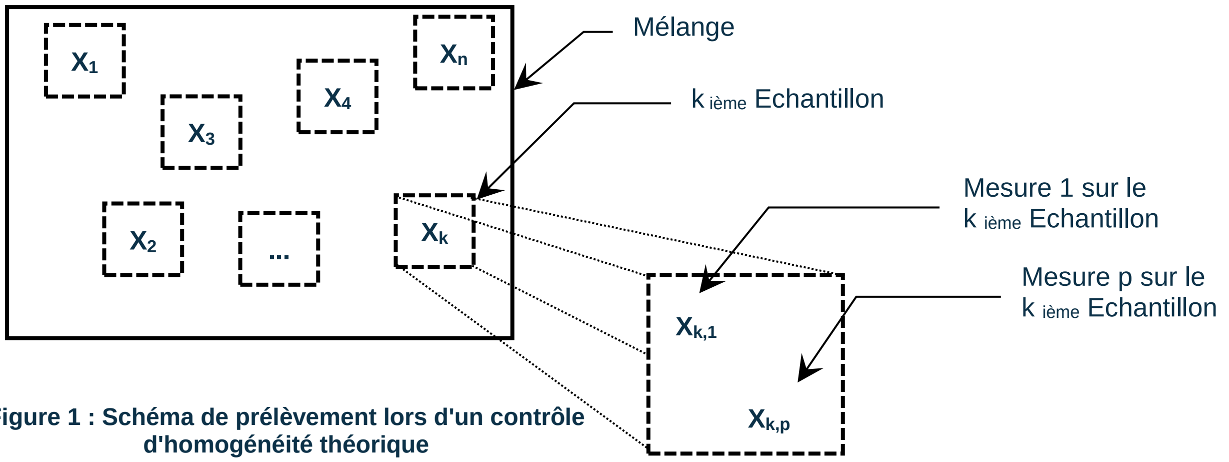

Figure 1 : Schéma de prélèvement lors d'un contrôle d'homogénéité théorique

Dans ce cadre, **pn** mesures sont obtenues. Le modèle aléatoire rend compte du fait que l'ensemble des échantillons collectés d'un mélange n'est qu'un groupe d'échantillons parmi d'autres qu'il aurait été possible de collecter. La réalisation d'une étude par modèle dit « fixe » ferait l'hypothèse que le groupe d'échantillons collectés est l'image parfaite de tous les groupes possibles et de la population entière elle-même. Le modèle aléatoire ne fait pas cette hypothèse et tente de tenir compte de l'erreur réalisée lors du choix des échantillons.
Les équations suivantes permettent de décrire avec précision les calculs réalisés. Dans ce modèle, l'évolution des mesures de concentration X_{ki} est donnée par :

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

ou

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

- μ = moyenne
- A_k est la différence entre la concentration de l'échantillon k et la concentration moyenne de la population μ. Ce terme rend compte de l'existence d'une variation de concentration entre chacun des échantillons. Chaque échantillon a sa propre valeur A_k. Il s'agit d'une variable aléatoire de moyenne 0 puisqu'elle est définie comme un écart à la moyenne μ. On suppose par simplicité que les A_k sont répartis selon une distribution normale de moyenne O et d' écart-type σ_A, notée $\mathcal{N}$ (0, σ_A).
- ε_{ki} rend compte :
  - des erreurs de mesures
  - de la variation de concentration entre prises d'essais dans un même échantillon.

Les A_k et les ε_{ki} sont supposés indépendants. On fait l'hypothèse que les ε_{ki} suivent une loi normale de moyenne nulle et d'écart type σ notée = $\mathcal{N}$ (0, σ).
Dans le cas du mélange, l'objectif va être d'estimer σ_A qui correspond alors à l'écart-type de la concentration du produit entre les échantillons pour conduire à une estimation correcte du coefficient de variation issu de l'opération d'homogénéisation.

L'analyse de variance conduit à la détermination des paramètres suivants :

- Moyenne totale μ

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

- La somme des carrés des écarts entre échantillons, SCE :

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

  SCE possède p-1 degrés de liberté soit un degré dans le cas de deux analyses par échantillon.

- La somme des carrés des écarts intra échantillons, SCI :

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

  SCI possède 2n-n degrés de liberté donc n degrés.

- La variance totale V_T :

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

- Le carré moyen entre échantillons ou variance inter-échantillons est estimé par CME :

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

- Le carré moyen intra échantillons ou variance intra-échantillon est estimé par CMI

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

Une comparaison entre la variance inter-échantillons (CME) et la variance intra-échantillon (CMI) est effectuée au moyen d'un test de Fischer. Cette comparaison est réalisée au moyen d'un $F_{calculé}$ qui peut être comparé à une table de Fischer-Snedecor (Snedecor et Cochran, 1971) en fonction du risque considéré :

$$F_{calculé} = \frac{CME}{CMI}$$

Le Tableau 1 donne quelques valeurs de $F_{théorique}$ en fonction du nombre d'échantillons pour deux analyses faites par échantillon.
Si le $F_{calculé}$ est supérieur au $F_{théorique}$, il est possible d'affirmer qu'il existe une différence significative entre les deux sources de variation. Si le Fcalculé est inférieur ou égal au Fthéorique, les deux sources de variation ne sont pas significativement différentes.

Nb d'échantillon	$F_{théorique}$
5	5.19
10	3.02
15	2.42
20	2.13
25	1.96
30	1.85

Tableau 1

L'existence d'une différence significative rend plus solide le calcul du CV qui en découle, car la méthode d'analyse utilisée est alors suffisamment précise pour observer les différences entre échantillons. A l'issue, trois variances peuvent être calculées : variance totale, variance résiduelle, variance homogénéité. Ces variances permettent le calcul des coefficients de variation correspondants.
La variance totale reste la même. Mais d'après le modèle aléatoire, le carré moyen intra échantillons est une estimation de la variance des

ε_{ki} :

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

et le carré moyen entre échantillons (CME) n'est pas une estimation directe de la variance des A_k, mais une estimation de la somme :

$$\mathbf{CME} = \sigma^{2} + \mathbf{p}\,\sigma_{A}^{2}$$

Ainsi, une estimation de la variance des A_k peut être obtenue par :

$$\boxed{\sigma_{A}^{2} = \frac{CME - CMI}{p} = V_{T} - CMI}$$

Le coefficient de variation entre échantillons dit « homogénéité » est obtenu par :

$$\boxed{\mathbf{CV}_{\mathbf{hom.}} = \mathbf{100}.\frac{\sqrt{\sigma_{A}^{2}}}{\mu}}$$

L'application du modèle aléatoire peut conduire à une variance résiduelle supérieure à la variance totale, ce qui a pour conséquence d'aboutir à une variance homogénéité (obtenue par différence des deux autres) négative. Dans ce cas, la variance et le coefficient de variation qui lui est lié sont fixés égaux à 0 par convention.
Ainsi, chaque étude de l'homogénéité d'un aliment (pour laquelle au moins deux analyses sont faites par échantillon) aboutit à la détermination de quatre paramètres :

- La moyenne totale.
- La signification de l'analyse de variance.
- Le $CV_{homogénéité}$ (Ecart-type du facteur échantillon ramené à la moyenne totale).
- Le CV_{total} (Ecart-type total ramené à la moyenne totale).

Le coefficient de variation a l'avantage d'être un nombre sans dimension, ce qui peut permettre à priori, de comparer des résultats entre eux. Cependant, il a des inconvénients qu'il convient de rappeler :

- **Il dépend de la moyenne** : si celle-ci diminue, à écart-type constant, le CV augmente. Ce paramètre doit donc toujours être vérifié, avant de poursuivre l'étude. Les contaminations croisées ont donc un impact potentiel sur l'homogénéité et inversement.
- **Il dépend de l'écart-type** : une détermination valide de cette valeur demande une population d'échantillons d'une taille supérieure à celle nécessaire à la détermination d'une moyenne fiable.
- **Il n'est pas sommable.** A la différence de la variance, il n'est pas possible de sommer des coefficients de variation. Cette seule différence conduit beaucoup de statisticiens à préférer la variance aux coefficients de variation comme indice de l'hétérogénéité.

2. Exemples

Soit un mélange dans lequel a été incorporé un traceur à un taux de 100 ppm. 10 échantillons sont prélevés dans le mélange et la concentration en traceur est analysée au moyen de deux prises d'essais dans chacun des échantillons. Ainsi, deux concentrations sont obtenues par échantillon (Tableau 2).

	Analyses	
Echantillons	**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

Tableau 2 : Exemple de résultats

La moyenne générale **μ** est à **97.3 ppm**. Elle est donc inférieure à celle attendue mais c'est elle qui sera utilisée pour le calcul du CV. On obtient également :

- SCE = **300.2**

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

Ainsi, le $F_{calculé}$ = 33.356/2.8 = **11.9** est supérieur au $F_{théorique}$ de **3.02** (obtenu pour un risque de 5 % et des degrés de liberté de 9 au numérateur et 10 au dénominateur). La variation entre les échantillons est donc significativement supérieure à celle intra échantillons.

La variance des A_k est donc :

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

et le coefficient de variation homogénéité est :

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

Le calcul de la variance totale permet l'obtention d'un coefficient de variation de 4.4 % et celui de la variance résiduelle (CMI) aboutit à un coefficient de 1.7 %. Ces résultats montrent bien la non-additivité des coefficients de variation. Ceci est un exemple très simple pour lequel l'analyse utilisée est de bonne qualité.

Dans l'exemple suivant (Tableau 3), la moyenne générale est toujours de 97.3 ppm.

	Analyses	
Echantillons	**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

Tableau 3 : Exemple de résultats

Cependant, les paramètres de calcul deviennent :

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

Le $F_{calculé}$ devient alors de 0.76 donc inférieur au $F_{théorique}$ de 3.02. Les variations intra et inter échantillons ne sont pas significativement différentes. La simple comparaison de CME et de CMI montre bien qu'elles sont du même ordre.
Il est possible de conclure que le coefficient de variation homogénéité est nul par rapport à celui engendré par la procédure analytique.
Si le calcul est poursuivi, la variance des A_k obtenue est négative ce qui est une particularité de ce modèle :

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

Dans ce cas, la variance totale peut être calculée (7.7 %) mais il convient de prendre en compte qu'elle est entièrement l'expression de variations engendrées par la procédure analytique.

3. Conclusion

L'application de ce modèle a pour avantage d'être intransigeant sur la qualité de la procédure analytique. Elle ne permet pas d'interprétation erronée de l'effet réel, de l'opération d'homogénéisation ou de celui d’un facteur stimulé qui pourrait conduire à des variations d’homogénéité (caractéristique physique d’un même additif, temps de mélange, etc. …). Il est également plus rigoureux du point de vue statistique.

4. Bibliographie

Cline A.L., 1978. Statistical evaluation of feed uniformity. Feedstuffs, 26 JUIN, 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. Sixième édition.