4. Méthode des Contrastes:

Lorsque le test associé à l'analyse de la variance aboutit à rejeter l'hypothèse de base d'égalité des traitements, il serait souhaitable de pouvoir définir avec plus de précision où se situent les différences. Diverses méthodes de comparaison utilisent l'idée de contrastes.

En général, un contraste est une combinaison linéaire des paramètres μ_i de la forme:

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i$$

$$pour\ laquelle\ \sum_{i=1}^{a} c_i = 0$$

En terme de contrastes, le corps d'hypothèses à tester s'expriment sous la forme:

$$H_0 : \sum_{i=1}^{a} c_i \mu_i = 0$$

$$H_a : \sum_{i=1}^{a} c_i \mu_i \neq 0$$

-- <u>Exemple:</u> Un ingénieur en développement de production s'intéresse à l'étude de la force extensible[1] d'une nouvelle fibre synthétique qui sera utilisée dans la confection de tissus destinés pour la production des T-shirts d'hommes. Suite à des expériences antérieures, cet ingénieur dispose de certaines informations sur l'étude,

- La force extensible est affectée par le poids exprimé en pourcentage du coton utilisé dans le mélange de matières formant cette fibre;
- L'augmentation de la teneur en coton intensifiera la force extensible, au moins initialement;
- Le contenu[2] du coton doit être rangé entre 10 et 40% (compte tenu de certaines caractéristiques de qualité possibles désirées au niveau du produit final).

Pour cela, l'ingénieur organise et exécute une expérience complètement randomisée avec 5 niveaux de pourcentages-poids de coton et 5 répliques pour tester si le pourcentage-poids de coton dans une fibre synthétique affecte réellement la force extensible. Les résultats des mesures sont consignés dans le tableau (1) ci-dessous:

[1] i.e. élastique.
[2] i.e la teneur en coton.

Tableau 1- Résultats de mesures.

Pourcentage-poids de coton	Force extensible observée (lb/in)				
	1	2	3	4	5
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

L'application de l'ANOVA rejette l'hypothèse de base où les traitements moyens diffèrent, i.e. le pourcentage du poids de coton dans la fibre synthétique affecte sensiblement la solidité moyenne de la force extensible. Cependant, on se demande lequel ou lesquels qui cause(nt) réellement cette différence? Au début de l'expérience, on peut soupçonner que les niveaux 4 et 5 des pourcentages-poids de cotons (i.e. 30% et 35%) produisent la même solidité de force extensible ce qui implique qu'on peut tester le corps d'hypothèse:

$$\begin{matrix} H_0: \mu_4 = \mu_5 \\ H_a: \mu_4 \neq \mu_5 \end{matrix} \quad (a)$$

Aussi, si on a soupçonné au début de l'expérience que les moyennes des faibles niveaux de pourcentages-poids de cotons (i.e. 1 et 2) ne diffèrent pas des moyennes des hauts niveaux de pourcentages-poids de cotons (i.e. 4 et 5) alors le corps d'hypothèse à tester sera:

$$\begin{matrix} H_0: \mu_1 + \mu_2 = \mu_4 + \mu_5 \\ H_a: \mu_1 + \mu_2 \neq \mu_4 + \mu_5 \end{matrix} \quad (b)$$

Alors au niveau du corps d'hypothèse (a), les contrastes seront définis par:

$$c_1 = c_2 = c_3 = 0, c_4 = 1 \; et \; c_5 = -1$$

Au niveau du corps d'hypothèse (b), les contrastes seront:

$$c_1 = c_2 = 1, c_3 = 0 \; et \; c_4 = c_5 = -1$$

-- <u>Remarque:</u> Un contraste sur les paramètres μ_i définit un contraste sur les paramètres α_i où on a:

$$\Gamma = \sum_{i=1}^{a} c_i\mu_i = \sum_{i=1}^{a} c_i(\mu + \alpha_i) = \sum_{i=1}^{a} c_i\alpha_i$$

$$pour\ laquelle\ \sum_{i=1}^{a} c_i = 0.$$

Deux méthodes existent pour tester l'idée de contrastes, à savoir l'approche du $t - test$ et l'approche du $F - test$.

i. <u>Approche du $t - test$:</u> En terme du total traitements, on suppose que le contraste d'intérêt est défini par:

$$C = \sum_{i=1}^{a} c_i y_{i.}$$

$$de\ variance\ V(C) = V\left(\sum_{i=1}^{a} c_i y_{i.}\right)$$

$$= V\left(\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}\, c_i\right)$$

$$= n\,\sigma^2 \sum_{i=1}^{a} c_i^2$$

$$lorsque\ les\ traitements\ ont\ tous\ le\ même\ nombre\ d'observations.$$

$$d'espérance\ E(C) = E\left(\sum_{i=1}^{a} c_i y_{i.}\right)$$

$$= n\sum_{i=1}^{a} c_i\, E\big(y_{ij}\big)$$

$$= n\sum_{i=1}^{a} c_i\, \mu_i$$

Si l'hypothèse $H_0\colon \sum_{i=1}^{a} c_i \mathrm{\mu_i} = 0$ est vraie alors le ratio

$$t_0 = \frac{\sum_{i=1}^{a} c_i\, y_{i.}}{\sqrt{n\,\sigma^2 \sum_{i=1}^{a} c_i^2}} \xrightarrow{s.H_0} (0,1) \qquad\qquad cas\ où\ \sigma^2\ est\ connu.$$

$Cas\ où\ \ \sigma^2\ est\ inconnu, estimé\ par\ CM_{SS_E} = \frac{SS_E}{N-a}$ où $s.H_0\colon \sum_{i=1}^{a} c_i \mathrm{\mu_i} = 0\ vraie, on\ a$

$$t_0 = \frac{\sum_{i=1}^{a} c_i\, y_{i.}}{\sqrt{n\,\hat{\sigma}^2\ \sum_{i=1}^{a} c_i^2}} = \frac{\sum_{i=1}^{a} c_i y_{i.}}{\sqrt{n\, CM_{SS_E}\ \sum_{i=1}^{a} c_i^2}} \rightsquigarrow \mathcal{T}(N-a)$$

$$Decision\colon\ rejet\ de\ H_0\ pour\ |t_0| > t_{\alpha/2,\ N-a}$$

ii. <u>Approche du $F - test$</u> : Par définition, on a

$$t_{\nu}^{2} = \mathcal{F}(1,\nu), alors$$

$$F_0 = t_0^2 = \left(\frac{\sum_{i=1}^{a} c_i y_{i.}}{\sqrt{n\, CM_{SS_E}\ \sum_{i=1}^{a} c_i^2}}\right)^2$$

$$= \frac{\left(\frac{\sum_{i=1}^{a} c_i y_{i.}}{\sqrt{n\ \sum_{i=1}^{a} c_i^2}}\right)^2}{CM_{SS_E}}$$

$$= \frac{SS_C/1}{CM_{SS_E}}$$

$$= \frac{CM_{SS_C}}{CM_{SS_E}} \rightsquigarrow \mathcal{F}(1, N-a)$$

$$Décision: rejet\ de\ l'hypothèse\ de\ base\ pour\ F_0 > F_{\alpha,(1,N-a)}.$$

$$où\ somme\ carrée\ des\ contrastes\ \ SS_C = \frac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{n\ \sum_{i=1}^{a} c_i^2} \rightsquigarrow \mathcal{X}^2(1).$$

iii. Intervalle de confiance d'un contraste: On considère le contraste d'intérêt suivant,

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i, \qquad pour\ lequel\ \sum_{i=1}^{a} c_i = 0$$

$$\Longrightarrow C = \sum_{i=1}^{a} c_i \bar{y}_{i.}$$

$$et\ V(C) = \frac{\sigma^2}{n} \sum_{i=1}^{a} c_i^2\ pour\ échantillons\ de\ tailles\ n_i\ égales.$$

$$\Longrightarrow IC_{\Gamma}^{(1-\alpha)\%} = \left[\sum_{i=1}^{a} c_i \bar{y}_{i.} \pm t_{\alpha/2,(N-a)} \sqrt{\frac{CM_{SS_E}}{n} \sum_{i=1}^{a} c_i^2}\right]$$

$$où\ CM_{SS_E}\ estimateur\ de\ \sigma^2\ inconnu.$$

iv. Contrastes Standardisés: Lorsqu'il existe plus qu'un contraste d'intérêt, il est utile de les évaluer tous sur la même échelle. Alors, on doit standardiser, normaliser le contraste de sorte qu'il présente une variance σ^2.

On définit le contraste standardisé par $\sum_{i=1}^{a} c_i^* y_i$ où $c_i^* = \frac{c_i}{\sqrt{n\ \sum_{i=1}^{a} c_i^2}}$.

v. <u>Sous échantillons de tailles inégales pour les traitements</u>: Certaines modifications à effectuer telles que:

* Définition de contraste exige que $\sum_{i=1}^{a} n_i c_i = 0$;

** Sous l'hypothèse de base $H_0: \sum_{i=1}^{a} c_i \mu_i = 0$, la statistique du test $t_0 = \frac{\sum_{i=1}^{a} c_i y_{i.}}{\sqrt{CM_{SS_E} \sum_{i=1}^{a} n_i c_i^2}}$.

*** Somme carrée des contrastes définie: $SS_C = \frac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{\sum_{i=1}^{a} n_i c_i^2}$

<u>A suivre.</u>