ÉQUATION D'ANALYSE DE LA VARIANCE (I, J)
(08 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'**équation d'analyse de la variance** est une équation classique de décomposition d'une **variance d'ensemble** en une somme de variances partielles, notamment pour effectuer un **test d'hypothèses** (cf aussi **analyse de la variance**, **décomposition de la variance**, **modèle d'analyse de la variance**, **modèle d'analyse de la covariance**, **modèle à variance composée**, **modèle linéaire**, **problème linéaire**, **théorème de COCHRAN**).

(i) Dans le cadre du **problème linéaire** général, on considère un **vecteur aléatoire** de carré intégrable y à valeurs dans l'**espace d'observation** $\mathbf{R}^N$ (muni d'une **structure** donnée) et une variété linéaire $\mathscr{L}$ de **rang** K de $\mathbf{R}^N$. Quelle que soit la **probabilité** P $\in \mathscr{P}$ (**famille** donnée), on suppose que $E_P\ y = \mu_P \in \mathscr{L}$ et que $V_P\ y = \Sigma_P$. Soit alors $\mathscr{M}$ une sous-variété de rang L de $\mathscr{L}$.

On considère un **problème de test** de l'**hypothèse de base**, souvent appelée **hypothèse nulle** (cf **test de l'hypothèse linéaire**) :

(1) $H_0 : \mu_P \in \mathscr{M}$

contre l'**alternative générale** :

(2) $H_a : \mu_P \in \mathscr{L}$.

Ici, l'hypothèse de base est une **hypothèse emboîtée** dans l'hypothèse alternative (ce qui peut se noter $\mathscr{M} \subset \mathscr{L}$ ou $\mathscr{M} \triangleleft \mathscr{L}$). On peut aussi noter H ($\mathscr{M}$) au lieu de H_0 et H ($\mathscr{L}$) au lieu de H_a . On considère alors la **projection oblique** (ie selon Σ_P) z ($\mathscr{M}$) de y sur $\mathscr{M}$.

Si $y \sim \mathscr{N}_N (\mu_P , \Sigma_P)$ (**loi normale multidimensionnelle**), on établit que :

(a) lorsque Σ_P est connue, la **statistique** (**forme quadratique**) :

(3) $Q_N = (z(\mathscr{L}) - z(\mathscr{M}))' \ \Sigma_P^{-1} \ (z(\mathscr{L}) - z(\mathscr{M}))$

suit, sous H_0 , la loi $\mathscr{X}^2$ (K-L) (**loi du chi-deux** à K - L **degrés de liberté**) : ie $Q_N \sim_{H0} \mathscr{X}^2$ (K-L).

Par suite, la **région critique** du test (rejet de H_a) est :

(4) $w = [Q_N > q_{1-\alpha}]$,

où $q_{1-\alpha}$ est le **quantile** d'ordre 1 - α de $\mathcal{X}^2$ (K-L) ;

(b) lorsque Σ_P n'est que partiellement connue (eg $\Sigma_P = \sigma_P^2 \,.\, \Omega_P$, où Ω_P est seule connue), la statistique :

(5) $$F_N = (K - L)^{-1} . (N - K) . \{Q (\mathscr{L}, \mathscr{M}) / Q (\mathbf{R}^N , \mathscr{L})\},$$

dans laquelle :

(6) $$\begin{aligned} Q (\mathscr{L}, \mathscr{M}) &= (z (\mathscr{L}) - z (\mathscr{M}))' \; \Omega_P^{-1} \; (z (\mathscr{L}) - z (\mathscr{M})), \\ Q (\mathbf{R}^N, \mathscr{L}) &= (y - z (\mathscr{L}))' \; \Omega_P^{-1} \; (y - z (\mathscr{L})), \end{aligned}$$

suit, sous H_0 , une loi $\mathscr{F}$ (K-L, N-K) (**loi de FISHER-SNEDECOR** à K - L et N - K degrés de liberté) : ie $F_N \sim \mathscr{F}$ (K-L, N-K).

La région critique du test (rejet de H_a) est :

(7) $$w = [F_N > q_{1-\alpha}],$$

où $q_{1-\alpha}$ est le quantile d'ordre 1 - α de $\mathscr{F}$ (K-L,N-K).

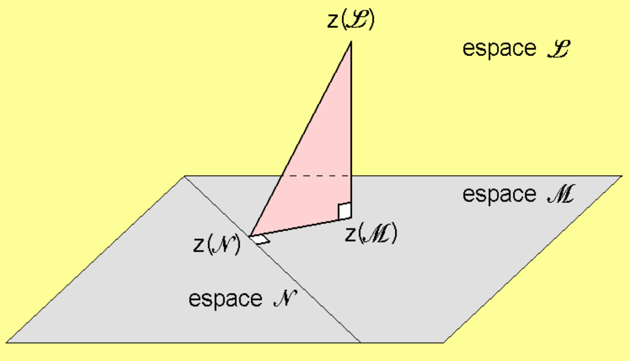
On appelle **tableau d'analyse de la variance** un tableau de la forme suivante (souvent utilisé dans l'étude d'un **plan d'expérience** et des modèles linéaires associés) :

Vecteur aléatoire	Degré de liberté	Variance « corrigée »
$y - z (\mathscr{L})$	N - K	$(N - K)^{-1}$ Q ($\mathbf{R}^N$, $\mathscr{L}$)
$z (\mathscr{L}) - z (\mathscr{M})$	K - L	$(K - L)^{-1}$ Q ($\mathscr{L}$, $\mathscr{M}$)
$y - z (\mathscr{M})$	N - L	$(N - L)^{-1}$ Q ($\mathbf{R}^N$, $\mathscr{M}$)

L'**équation d'analyse de la variance** associée s'écrit :

(8) $$Q (\mathbf{R}^N, \mathscr{M}) = Q (\mathbf{R}^N, \mathscr{L}) + Q (\mathscr{L}, \mathscr{M}).$$

(ii) Cette démarche s'étend directement au test d'une **suite d'hypothèses emboîtées**. Ainsi, dans le cas où trois hypothèses H ($\mathscr{L}$), H ($\mathscr{M}$) et H ($\mathscr{N}$) sont resp associées à des variétés linéaires $\mathscr{L}$, $\mathscr{M}$ et $\mathscr{N}$, de dimensions resp K, L et M et tq $\mathscr{N} \subset \mathscr{M} \subset \mathscr{L}$, on note resp z ($\mathscr{L}$), z ($\mathscr{M}$) et z ($\mathscr{N}$) les projections obliques (ie selon Σ_P) de y sur $\mathscr{L}$, $\mathscr{M}$ et $\mathscr{N}$ (cf graphique ci-dessous).



L'**équation d'analyse de la variance** s'écrit alors :

(9) $$Q(\mathbf{R}^N, \mathscr{N}) = Q(\mathbf{R}^N, \mathscr{L}) + Q(\mathscr{L}, \mathscr{M}) + Q(\mathscr{M}, \mathscr{N}),$$

et le tableau d'analyse de la variance correspondant s'écrit :

Vecteur aléatoire	Degré de liberté	Variance « corrigée »
$y - z(\mathscr{L})$	N - K	$(N - K)^{-1} Q(\mathbf{R}^N, \mathscr{L})$
$z(\mathscr{L}) - z(\mathscr{M})$	K - L	$(K - L)^{-1} Q(\mathscr{L}, \mathscr{M})$
$z(\mathscr{M}) - z(\mathscr{N})$	L - M	$(L - M)^{-1} Q(\mathscr{M}, \mathscr{N})$
$y - z(\mathscr{N})$	N - M	$(N - M)^{-1} Q(\mathbf{R}^N, \mathscr{N})$