# MAT7381 Chapitre 3
# Le modèle linéaire général

## Définition du modèle

L'équation du modèle linéaire général est

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

où $\boldsymbol{y}$ est un vecteur $n \times 1$ d'observations aléatoires, $\mathbf{X}$ est une matrice $n \times q$ de constantes connues, $\boldsymbol{\beta}$ est un vecteur $q \times 1$ de paramètres inconnus et $\boldsymbol{\varepsilon}$ est un vecteur $n \times 1$ d'erreurs inobservables. On suppose généralement que

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ et } \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$$

Finalement, nous supposons que $\mathbf{X}$ est de rang plein $q$.

Ces suppositions sont suffisantes pour obtenir des estimateurs ponctuels sans biais. Pour développer de procédures de test et des intervalles de confiance, nous ajouterons l'hypothèse que $\boldsymbol{y}$ est de loi normale multivariée.

## Estimateurs

Pour déterminer un estimateur pour le vecteur $\boldsymbol{\beta}$, nous utilisons le principe des moindres carrés, qui s'énonce comme ceci: l'estimateur des moindres carrés est la valeur $\hat{\boldsymbol{\beta}}$ du vecteur $\boldsymbol{\beta}$ qui minimise

$$\mathrm{Q} = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})'(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})$$

Il s'agit donc de trouver un vecteur $\mathbf{X}\hat{\boldsymbol{\beta}} \in \mathcal{C}(\mathbf{X})$, pour lequel la longueur (au carré) du vecteur $\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ est minimale. Il est géométriquement clair que ce vecteur $\mathbf{X}$ doit être la projection orthogonale de $\boldsymbol{y}$ sur $\mathcal{C}(\mathbf{X})$, c'est-à-dire,

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y},$$

l'existence de $(\mathbf{X}'\mathbf{X})^{-1}$ étant assurée par le fait que $\mathbf{X}$ est de rang plein. Multipliant les deux membres de l'équation par $\mathbf{X}'$ à gauche on a

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \mathbf{X}'\boldsymbol{y}$$

ce qui entraîne

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$$

Une démonstration plus formelle est également possible. Nous avons

$$\begin{aligned}\mathrm{Q} &= [\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}]'[\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}] \\ &= [\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}]'[\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}].\end{aligned}$$

Mais

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \mathbf{P}\boldsymbol{y} \quad \text{et} \quad \boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y} - \mathbf{P}\boldsymbol{y} = (\mathbf{I} - \mathbf{P})\boldsymbol{y},$$

où $\mathbf{P}$ est le projecteur orthogonal sur $\mathcal{C}(\mathbf{X})$, soit

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

$\mathbf{P}$ et $\mathbf{I} - \mathbf{P}$ sont symétriques et idempotentes. Nous avons donc

$$\mathrm{Q} = [(\mathbf{I} - \mathbf{P})\boldsymbol{y} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]'[(\mathbf{I} - \mathbf{P})\boldsymbol{y} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]$$

$= \boldsymbol{y}'(\mathbf{I} - \mathbf{P})'(\mathbf{I} - \mathbf{P})\boldsymbol{y} + (\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta) + 2\boldsymbol{y}'(\mathbf{I}-\mathbf{P})'\mathbf{X}(\hat{\beta} - \beta)$ .

Mais

$\boldsymbol{y}'(\mathbf{I} - \mathbf{P})'(\mathbf{I} - \mathbf{P})\boldsymbol{y} = \boldsymbol{y}'(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P})\boldsymbol{y} = \boldsymbol{y}'(\mathbf{I} - \mathbf{P})\boldsymbol{y}$, puisque $(\mathbf{I} - \mathbf{P})$ est symétrique et idempotente.

Et $(\mathbf{I}-\mathbf{P})'\mathbf{X} = (\mathbf{I}-\mathbf{P})\mathbf{X} = \mathbf{X}-\mathbf{PX} = \mathbf{X}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}-\mathbf{X} = \mathbf{0}$.

Donc $Q = \boldsymbol{y}'(\mathbf{I} - \mathbf{P})\boldsymbol{y} + (\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)$.

Étant donné que le premier terme, $\boldsymbol{y}'(\mathbf{I}-\mathbf{P})\boldsymbol{y}$, ne dépend pas de $\beta$, il s'agit de minimiser le deuxième, $(\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)$. Rappelons qu'une matrice de la forme $\mathbf{X}'\mathbf{X}$ est toujours semi définie positive, et puisque $\mathbf{X}$ est de rang plein, $\mathbf{X}'\mathbf{X}$ est en fait *définie* positive. Donc $(\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta) > 0$ pour tout $\beta \neq \hat{\beta}$, et par conséquent $(\hat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)$ atteint son minimum de 0 si et seulement si $\beta = \hat{\beta}$ .

**Distribution des estimateurs**

1. $\hat{\beta}$ est un vecteur normal, puisqu'il est fonction linéaire d'un vecteur normal. Son espérance est

$$E(\hat{\beta}) = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta$$

   c'est-à-dire, $\hat{\beta}$ est un estimateur sans biais de $\beta$. Sa variance est

$$\mathrm{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{Var}(\boldsymbol{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

2. La somme des carrés résiduelle est

$$SCR = (\boldsymbol{y} - \mathbf{X}\hat{\beta})'(\boldsymbol{y} - \mathbf{X}\hat{\beta}) = [(\mathbf{I} - \mathbf{P})\boldsymbol{y}]'[(\mathbf{I} - \mathbf{P})\boldsymbol{y}] = \boldsymbol{y}'(\mathbf{I} - \mathbf{P})\boldsymbol{y}$$

   grâce à l'idempotence et la symétrie de $\mathbf{P}$. Il découle immédiatement des théorèmes sur les distributions de formes quadratiques que

$$\frac{\boldsymbol{y}'(\mathbf{I}-\mathbf{P})\boldsymbol{y}}{\sigma^2} \sim \chi^2_{n-q}$$

   Le paramètre de non-centralité $\lambda$ est nul car si $E(\boldsymbol{y}) = \boldsymbol{\mu} = \mathbf{X}\beta$, $\lambda = \sigma^2\boldsymbol{\mu}'(\mathbf{I}-\mathbf{P})\boldsymbol{\mu} = 0$, puisque $\mathbf{PX} = \mathbf{X}$ et donc $(\mathbf{I}-\mathbf{P})\boldsymbol{\mu} = (\mathbf{I}-\mathbf{P})\mathbf{X}\beta = \mathbf{0}$.

3. On déduit aisément de 2 ci-dessus que $\hat{\sigma}^2 = SCR/(n-q)$ est un estimateur sans biais de $\sigma^2$.

4. Les vecteurs $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$ et $(\mathbf{I} - \mathbf{P})\boldsymbol{y}$ sont indépendants puisque $\mathrm{Cov}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} ; (\mathbf{I}-\mathbf{P})\boldsymbol{y}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{P})' = \mathbf{0}$. Donc $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants.

5. Si $\ell$ est un vecteur $q \times 1$, la variable $\ell'\hat{\beta}$ suit une loi $\mathcal{N}[\ell'\beta ; \sigma^2_{\ell'\hat{\beta}}]$ où $\sigma^2_{\ell'\hat{\beta}} = \sigma^2\ell'(\mathbf{X}'\mathbf{X})^{-1}\ell$. On a donc

$$\frac{\ell'\hat{\beta} - \ell'\beta}{\sigma_{\ell'\hat{\beta}}} \sim \mathcal{N}(0 ;1)$$

   Les résultats précédents nous permettent ensuite de conclure que

$$\frac{\ell'\hat{\beta} - \ell'\beta}{\hat{\sigma}_{\ell'\hat{\beta}}} \sim t_{n-q}$$

   où $\hat{\sigma}_{\ell'\hat{\beta}} = \hat{\sigma}\sqrt{\ell'(\mathbf{X}'\mathbf{X})^{-1}\ell}$

6. Les résultats en 5 ont un analogue matriciel. Si $\mathbf{L}$ est une matrice de constantes $q \times r$ de rang $r \times q$, le vecteur aléatoire $\mathbf{L}'\hat{\boldsymbol\beta}$ suit une loi $\mathcal{N}_r[\mathbf{L}'\boldsymbol\beta\ ;\ \sigma^2\mathbf{L}'(\mathbf{X}'\mathbf{X})\text{-}1\mathbf{L}]$. Donc

$$\frac{(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)'[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)/r}{\sigma^2} \sim \chi^2_r$$

De là, on peut déduire que la variable

$$\frac{(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)'[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)/r}{\hat\sigma^2} \sim \mathcal{F}_{r,n-q}.$$

**Intervalles de confiance**

Si $\ell$ est un vecteur $q \times 1$, on peut déterminer un intervalle de confiance pour $\ell'\boldsymbol\beta$ en utilisant le résultat 5 ci-dessus. On obtient

$$\ell'\hat{\boldsymbol\beta} - t_{n-q,\alpha/2}\,\hat\sigma_{\ell'\hat{\boldsymbol\beta}} \leq \ell'\boldsymbol\beta \leq \ell'\hat{\boldsymbol\beta} + t_{n-q,\alpha/2}\,\hat\sigma_{\ell'\hat{\boldsymbol\beta}}$$

où $t_{n-q;\sigma/2}$ est le point critique correspondant à une loi de Student à $n - q$ degrés de liberté et $\hat\sigma_{\ell'\hat{\boldsymbol\beta}} = \hat\sigma\,[\ell'(\mathbf{X}'\mathbf{X})^{-1}\ell]^{1/2}$ est l'estimateur de l'écart-type de $\ell'\boldsymbol\beta$**.** Ainsi, en prenant pour $\ell'$ les vecteurs (1 ; 0 ; 0 ; ... ; 0) ; (0 ; 1 ; 0 ; ... ; 0) ; ... ; (0 ; 0 ; 0 ; ... ; 1), nous pouvons déterminer un intervalle de confiance pour chacune des composantes de $\boldsymbol\beta$.

**Tests d'hypothèses linéaires**

Une hypothèse linéaire générale est la suivante :

$$H_o\text{:}\ \ \mathbf{L}'\boldsymbol\beta = \boldsymbol\delta$$

où $\mathbf{L}$ est une matrice $q \times r$ ($r \leq q$) de rang plein et $\boldsymbol\delta$ est un vecteur $r \times 1$ de constantes. Il est évident que le test d'une telle hypothèse devra être basé sur le vecteur aléatoire $\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta$ dont l'espérance $\mathbf{L}'\boldsymbol\beta - \boldsymbol\delta$ est nulle si et seulement si $H_o$ est vraie. Donc la norme de $\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta$ devrait être petite si $H_o$ est vraie et grande sinon. La norme la plus naturelle, pour un vecteur normal, est définie par

$$|\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta\|^2 = (\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)'[\mathrm{Var}(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)]^{-1}(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)$$

$$= (\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)'[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)/\sigma^2$$

Cette forme quadratique suit une loi $\chi^2_r(\lambda)$ avec

$$\lambda = (\mathbf{L}'\boldsymbol\beta - \boldsymbol\delta)'[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}(\mathbf{L}'\boldsymbol\beta - \boldsymbol\delta)/\sigma^2$$

Elle est donc centrale si et seulement si $H_o$ est vraie. Puisque $\sigma^2$ n'est pas connue, nous la remplaçons par $\hat\sigma^2$ et nous divisons le numérateur par le nombre de degrés de liberté $r$. Nous obtenons le rapport

$$\frac{(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)'[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}(\mathbf{L}'\hat{\boldsymbol\beta} - \boldsymbol\delta)/r}{\hat\sigma^2}$$

qui suit une loi $F$ centrale à $r$ et $n - q$ degrés de liberté si et seulement si $H_o$ est vraie.

**Différences de sommes de carrés résiduelles**

Nous montrons ici que le numérateur de la statistique $F$ peut s'exprimer comme une différence de deux sommes de carrés résiduelles $SCR_o$ - SCR, où SCR est la somme des carrés résiduelle sous le modèle et $SCR_o$ est la somme des carrés résiduelle *sous le modèle réduit,* c'est-à-dire, sous $H_o$.

*D'abord, le cas où* $\boldsymbol{\delta} = \mathbf{0}$.

**Cas $\boldsymbol{\delta} = \mathbf{0}$**. Nous considérons donc l'hypothèse

$$H_o\text{: } \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$$

La somme des carrés résiduelle sous le modèle est

$$\text{SCR} = \boldsymbol{y}'(\mathbf{I} - \mathbf{P})\boldsymbol{y} \text{ où } \mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Nous devons déterminer maintenant $\text{SCR}_o$, la somme des carrés résiduelle sous l'hypothèse nulle. L'hypothèse $\mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ est équivalente à

$$H_o\text{: } \boldsymbol{\beta} \in \mathcal{N}(\mathbf{L}')$$

Nous supposons sans perte de généralité que $\mathbf{L}$ est de rang plein $r$, auquel cas la dimension de $\mathcal{N}(\mathbf{L}')$ est $q$-$r$. Soit $\mathbf{T}$ une matrice $q \times (q - r)$ dont les colonnes constituent une base pour $\mathcal{N}(\mathbf{L}')$. C'est-à-dire,

$$\mathcal{C}(\mathbf{T}) = \mathcal{N}(\mathbf{L}')$$

L'hypothèse nulle $\mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ est donc équivalente à

$$H_o\text{: } \boldsymbol{\beta} = \mathbf{T}\boldsymbol{\alpha}$$

Le modèle réduit est donc

$$\boldsymbol{y} = \mathbf{X}\mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

Par conséquent, la somme des carrés résiduelle pour le modèle réduit est

$$\text{SCR}_o = \boldsymbol{y}'(\mathbf{I} - \mathbf{P_o})\boldsymbol{y}$$

où

$$\mathbf{P}_o = \mathbf{X}\mathbf{T}(\mathbf{T}'\mathbf{X}'\mathbf{X}\mathbf{T})^{-1}\mathbf{T}'\mathbf{X}'$$

Nous avons donc

$$\text{SCR}_o - \text{SCR} = \boldsymbol{y}'(\mathbf{P} - \mathbf{P}_o)\boldsymbol{y}$$

Nous devons démontrer que cette forme quadratique est identique à celle présentée plus haut, soit

$$[\mathbf{L}'\hat{\boldsymbol{\beta}}]'[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}} = \boldsymbol{y}'\mathbf{P}_1\boldsymbol{y},$$

où

$$\mathbf{P}_1 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

c'est-à-dire, il faut montrer que

$$\mathbf{P} - \mathbf{P}_o = \mathbf{P}_1$$

Étant donné que $\mathbf{P}$-$\mathbf{P}_o$ et $\mathbf{P}_1$ sont des projecteurs orthogonaux, il suffira de montrer que leurs colonnes engendrent le même espace. Pour montrer ceci, on montre que $\mathcal{C}(\mathbf{P}_1) \subset \mathcal{C}(\mathbf{P}\text{-}\mathbf{P}_o)$ et ensuite que $r(\mathbf{P}\text{-}\mathbf{P}_o) = r(\mathbf{P}_1)$.

Il est clair que $\mathbf{P}\mathbf{P}_1 = \mathbf{P}_1$, puisque $\mathbf{P}\mathbf{X} = \mathbf{X}$, et on vérifie aisément $\mathbf{P}_o\mathbf{P}_1 = \mathbf{0}$, en utilisant le fait que $\mathbf{T}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L} = \mathbf{T}'\mathbf{L} = \mathbf{0}$. Donc $(\mathbf{P}\text{-}\mathbf{P}_o)\mathbf{P}_1 = \mathbf{P}_1$, ce qui entraîne que $\mathcal{C}(\mathbf{P}_1) \subset \mathcal{C}(\mathbf{P}\text{-}\mathbf{P}_o)$.

Par l'idempotence de $\mathbf{P}$-$\mathbf{P}_o$ et $\mathbf{P}_1$, $r(\mathbf{P}\text{-}\mathbf{P}_o) = r(\mathbf{P}_1)$ si et seulement si $tr(\mathbf{P}\text{-}\mathbf{P}_o) = tr(\mathbf{P}_1)$. Or

$tr(\mathbf{P}\text{-}\mathbf{P}_o) = tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}\mathbf{T}(\mathbf{T}'\mathbf{X}'\mathbf{X}\mathbf{T})^{-1}\mathbf{T}'\mathbf{X}') = tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} - (\mathbf{T}'\mathbf{X}'\mathbf{X}\mathbf{T})^{-1}\mathbf{T}'\mathbf{X}'\mathbf{X}\mathbf{T}) = q - (q\text{-}r) = r$

et

$tr(\mathbf{P}_1) = tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$

sont des projecteurs orthogonaux sur le même espace et sont donc égaux par l'unicité des projecteurs orthogonaux.

Ceci complète la démonstration pour le cas où $\boldsymbol{\delta} = \mathbf{0}$.

**Cas $\delta \neq 0$.** Pour passer au cas plus général, nous devons reformuler $H_o$. Si $\mathbf{L}'\boldsymbol{\beta} = \boldsymbol{\delta}$, alors $\boldsymbol{\beta} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\gamma}$, où **T** est définie comme ci-dessus [une matrice $q \times (q - r)$ dont les colonnes sont une base de $\mathcal{N}(\mathbf{L}')$] et $\boldsymbol{\gamma}$ satisfait $\mathbf{L}'\boldsymbol{\gamma} = \boldsymbol{\delta}$. Le modèle réduit devient

$$\boldsymbol{y} = \mathbf{XT}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

ou encore

$$\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma} = \mathbf{XT}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

et la somme des carrés résiduelle sous cette hypothèse est

$$(\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})'(\mathbf{I}-\mathbf{P}_o)(\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})$$

Puisque la somme des carrés résiduelle $\boldsymbol{y}'(\mathbf{I} - \mathbf{P})\boldsymbol{y}$ sous le modèle peut aussi s'écrire sous la forme

$$\text{SCR} = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})'(\mathbf{I}-\mathbf{P})(\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})$$

(le vecteur **XT** étant annulé par $\mathbf{I} - \mathbf{P}$), la différence $\text{SCR}_o$ - SCR s'écrit

$$\text{SCR}_o - \text{SCR} = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})'(\mathbf{P}-\mathbf{P}_o)(\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})]$$

Cette forme quadratique doit être comparée au numérateur

$$(\mathbf{L}'\hat{\boldsymbol{\beta}} - \boldsymbol{\delta})'[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}(\mathbf{L}'\hat{\boldsymbol{\beta}} - \boldsymbol{\delta})$$

de la statistique *F*. Le facteur $\mathbf{L}'\hat{\boldsymbol{\beta}} - \boldsymbol{\delta}$ qui y apparaît peut s'écrire comme

$$\begin{aligned}\mathbf{L}'\hat{\boldsymbol{\beta}} - \boldsymbol{\delta} &= \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} - \mathbf{L}'\boldsymbol{\gamma} \\ &= \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} - \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\gamma} \\ &= \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma}),\end{aligned}$$

et la forme quadratique au numérateur est donc

$$(\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma}) = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})'\mathbf{P}_1(\boldsymbol{y} - \mathbf{X}\boldsymbol{\gamma})$$

Ceci complète la démonstration, puisque nous avons déjà démontré que les matrices $\mathbf{P} - \mathbf{P}_o$ et $\mathbf{P}_1$ sont identiques.

### Le théorème de Gauss-Markov

Nous avons obtenu l'estimateur $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$ par la méthode des moindres carrés, qui n'est qu'un critère parmi d'autres pour choisir un estimateur. Les autres critères mènent-ils au même estimateur? La réponse est oui, la plupart des critères connus mènent à l'estimateur $\hat{\boldsymbol{\beta}}$. L'un est la méthode du maximum de vraisemblance : si le vecteur $\boldsymbol{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} ; \sigma^2\mathbf{I})$, la valeur de $\boldsymbol{\beta}$ qui maximise la fonction de vraisemblance est $\hat{\boldsymbol{\beta}}$. Un autre critère consiste à choisir, dans une classe restreinte d'estimateurs, le « meilleur » estimateur. Une façon de restreindre la classe des estimateurs est d'exiger un estimateur *linéaire*. On restreint la classe encore plus en exigeant un estimateur sans biais. Il se peut que dans cette classe restreinte, il existe un seul « meilleur » estimateur. Quel sens donne-t-on à « meilleur »? Habituellement, lorsqu'il s'agit d'un estimateur scalaire, un estimateur est le « meilleur » s'il a la plus petite variance. Mais lorsque l'estimateur est un vecteur, la notion de « meilleur » se généralise difficilement. Le cas présent, cependant, est un des cas relativement rares où la généralisation se fait de façon particulièrement satisfaisante. L'énoncé exact est dans le *théorème de Gauss-Markov*.

*Théorème de Gauss-Markov***.** *Soit* $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$ *l'estimateur des moindres carrés. Soit* $\tilde{\boldsymbol{\beta}} = \mathbf{A}\boldsymbol{y}$ *un autre estimateur linéaire sans biais pour* $\boldsymbol{\beta}$. *Alors*

$$\text{Var}(\tilde{\boldsymbol{\beta}}) - \text{Var}(\hat{\boldsymbol{\beta}}) \textit{ est semi-définie positive.}$$

*Démonstration* $\tilde{\beta}$ est sans biais $\Rightarrow$ E($\mathbf{A}\boldsymbol{y}$) = $\beta$ $\forall\beta$.

Puisque

$$E(\mathbf{A}\boldsymbol{y}) = \mathbf{A}E(\boldsymbol{y}) = \mathbf{A}\mathbf{X}\beta,$$

alors

$$\mathbf{A}\mathbf{X}\beta = \beta \quad \forall\beta \Rightarrow \mathbf{A}\mathbf{X} = \mathbf{I} \ (\Rightarrow \mathbf{X}'\mathbf{A}' = \mathbf{I}).$$

Or $\mathrm{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ et $\mathrm{Var}(\tilde{\beta}) = \sigma^2\mathbf{A}\mathbf{A}'$, d'où

$$\begin{aligned}\mathrm{Var}(\tilde{\beta}) - \mathrm{Var}(\hat{\beta}) &= \sigma^2\mathbf{A}\mathbf{A}' - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\\ &= \sigma^2\mathbf{A}\mathbf{A}' - \sigma^2\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}'\\ &= \sigma^2\mathbf{A}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{A}'\\ &= \sigma^2\mathbf{A}(\mathbf{I} - \mathbf{P})\mathbf{A}'\end{aligned}$$

où $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Puisque $\mathbf{I} - \mathbf{P}$ est symétrique et idempotente, $\sigma^2\mathbf{A}(\mathbf{I} - \mathbf{P})\mathbf{A}' = \sigma^2\mathbf{A}(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P})'\mathbf{A}'$ , ce qui est clairement semi définie positive. ■

Ce théorème montre que toute fonction linéaire de $\hat{\beta}$ est moins dispersée que la même fonction linéaire de $\tilde{\beta}$ : si $\ell$ est un vecteur de constantes,

$$\mathrm{Var}(\ell'\tilde{\beta}) - \mathrm{Var}(\ell'\hat{\beta}) = \ell'\mathrm{Var}(\tilde{\beta})\ell - \ell'\mathrm{Var}(\hat{\beta})\ell = \ell'\sigma^2[\mathrm{Var}(\tilde{\beta}) - \mathrm{Var}(\hat{\beta})]\ell \geq 0$$

quelle que soit la valeur de $\ell$, parce que, par le théorème de Gauss-Markov, $[\mathrm{Var}(\tilde{\beta}) - \mathrm{Var}(\hat{\beta})]$ est semi définie positive.

### Table d'analyse de variance

Une *analyse de varianc*e est une décomposition d'une somme de carrés en une somme de sommes de carrés. La principale décomposition provient de la projection de $\boldsymbol{y}$ sur deux espaces orthogonaux, soit

$$\boldsymbol{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} + [\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{y}$$

Le premier terme est $\mathbf{X}\hat{\boldsymbol{\beta}}$ et fournit une estimation de $\beta$ étant donnée que son espérance est $\mathbf{X}\beta$. Le deuxième est de moyenne nulle et permet d'estimer la *variance* car son espérance est $(n\text{-}q)\sigma^2$, indépendante de $\beta$. Par l'orthogonalité des espaces $\mathcal{C}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ et $\mathcal{C}([\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'])$ on a la décomposition

$$\boldsymbol{y}'\boldsymbol{y} = \boldsymbol{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} + \boldsymbol{y}'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{y} = \hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{X})^{-1}\hat{\boldsymbol{\beta}} + \boldsymbol{y}'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{y}$$

La première somme, $\hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{X})^{-1}\hat{\boldsymbol{\beta}}$, aura tendance à prendre de petites valeurs lorsque $\boldsymbol{\beta} = \mathbf{0}$. La somme des carrés $\boldsymbol{y}'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{y}$ est appelé *somme des carrés résiduelle*.

Les logiciels statistiques présentent automatiquement une décomposition de la variance de $y_1, \ldots , y_n$ en deux sommes de carrés sous la forme d'une table appelée *analyse de variance*, une décomposition qui, typiquement, n'est pas tout à fait celle qu'on vient de définir. Par défaut, ils supposent que l'une des colonnes de $\mathbf{X}$ est une colonne de « 1 », étant donnée que dans la plupart des applications. Supposons donc que la première colonne de la matrice $\mathbf{X}$ est une colonne de « 1 », qu'on désignera par $\boldsymbol{e}$. Donc $\mathbf{X}$ = $[\boldsymbol{e} \mid \mathbf{X}_1]$ et $\boldsymbol{\beta} = \begin{bmatrix}\beta_o\\ \boldsymbol{\beta}_1\end{bmatrix}$. La somme des carrés qui est décomposée est $\boldsymbol{y}'[\mathbf{I} - \boldsymbol{e}\boldsymbol{e}'/n]\boldsymbol{y} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ , la dispersion des $y_i$ par rapport à leur moyenne $\bar{y}$ . Elle correspond à la somme des carrés résiduelle dans un modèle dans lequel les observations sont toutes de même moyenne et donc ne dépendent pas des variables exogènes. La somme des carrées expliquée est $\boldsymbol{y}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \boldsymbol{ee}'/n]\boldsymbol{y}$ et non $\boldsymbol{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$. Nous obtenons le tableau suivant :

| Source | Somme des carrés | Degrés de liberté | Moyenne des carrés | Statistique $F$ | Valeur de $p$ |
|---|---|---|---|---|---|
| Expliquée | SCE = $\boldsymbol{y}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'-\boldsymbol{ee}'/n]\boldsymbol{y}$ | $q$-1 | MCE = SCE/($q$-1) | $F = \dfrac{\text{MCE}}{\text{MCR}}$ | $P(F_{q-1;n-q} \geq F)$ |
| Résiduelle | SCR = $\boldsymbol{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{y}$ | $n$-$q$ | MCR = SCR/($n$-$q$) | | |
| Totale | SCT = $\boldsymbol{y}'[\mathbf{I} - \boldsymbol{ee}'/n]\boldsymbol{y}$ | $n$-1 | | | |

Ce que cette table permet de faire, c'est tester une hypothèse globale concernant l'effet des variables exogènes sur la variable dépendante. La somme de carrées *expliquée*, SCE, est une différence de deux sommes de carrés résiduelles,

$$\text{SCE} = \boldsymbol{y}'[\mathbf{I} - \boldsymbol{ee}'/n]\boldsymbol{y} - \boldsymbol{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{y},$$

qui compare le modèle $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ au modèle restreint $\boldsymbol{y} = \beta_o\boldsymbol{e} + \boldsymbol{\varepsilon}$. Donc SCE est le numérateur d'une statistique $F$ pour tester l'hypothèse que $\boldsymbol{\beta}_1 = \mathbf{0}$. Il est instructif, d'ailleurs, de noter que SCE = $\hat{\boldsymbol{\beta}}_1'(\mathbf{X}'\mathbf{C}\mathbf{X})^{-1}\hat{\boldsymbol{\beta}}_1$, où $\mathbf{C} = \mathbf{I}-\boldsymbol{ee}'/n$. On vérifie ceci en montrant que $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \boldsymbol{ee}'/n + \mathbf{C}\mathbf{X}_1(\mathbf{X}_1'\mathbf{C}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{C}$.

### L'effet des dépendances entre les variables exogènes

Des dépendances entre les variables exogènes tendent à diminuer l'efficacité des estimations et à réduire la puissance des tests. Considérons les partitions conformes $\mathbf{X} = [\mathbf{X}_1 \mid \mathbf{X}_2]$ et $\boldsymbol{\beta} = \begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix}$. L'estimateur de $\boldsymbol{\beta}$ est

$$\hat{\boldsymbol{\beta}} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \begin{bmatrix}\hat{\boldsymbol{\beta}}_1\\ \hat{\boldsymbol{\beta}}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1'\\ \mathbf{X}_2'\end{bmatrix}\boldsymbol{y}$$

et la matrice de covariance de $\hat{\boldsymbol{\beta}}$ est

$$\sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{11} & \left(\mathbf{X}_1'\mathbf{X}_2\right)^{12}\\ \left(\mathbf{X}_2'\mathbf{X}_1\right)^{21} & \left(\mathbf{X}_2'\mathbf{X}_2\right)^{22}\end{bmatrix},$$

où

$$(\mathbf{X}_1'\mathbf{X}_1)^{11} = (\mathbf{X}_1'\mathbf{X}_1)^{-1} + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2[\mathbf{X}_2'(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_2]^{-1}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}$$

Donc

$$\text{Var}(\hat{\boldsymbol{\beta}}_1) = \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{11} = \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1} + \sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2[\mathbf{X}_2'(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_2]^{-1}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}$$

Si les colonnes de $\mathbf{X}_2$ avaient été orthogonales à celles de $\mathbf{X}_1$, la 2$^e$ matrice dans la somme ci-dessus aurait été nulle. Autrement, elle est définie positive et vient donc « augmenter » la variance de $\hat{\boldsymbol{\beta}}_1$.

Le test de l'hypothèse $H_o$ : $\boldsymbol{\beta}_1 = \mathbf{0}$ perd de sa puissance lorsque les variables exogènes sont corrélées. Le numérateur de la statistique $F$ pour tester $H_o$ est $\boldsymbol{y}[\mathbf{P} - \mathbf{P}_2]\boldsymbol{y}$, où $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ et $\mathbf{P}_2 = \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$. L'espérance de cette forme quadratique est

$$\boldsymbol{\beta}'\mathbf{X}'[\mathbf{P} - \mathbf{P}_2]\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}_1'\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1'\mathbf{X}_1'\mathbf{X}_1\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1'\mathbf{X}_1'\mathbf{P}_2\mathbf{X}_1\boldsymbol{\beta}_1 .$$

Le deuxième terme vient réduire l'espérance du numérateur et donc réduit la puissance. Il disparaît si les colonnes de $\mathbf{X}_1$ et celles de $\mathbf{X}_2$ sont orthogonales.

### Remarque concernant la précision de l'estimation

On sait que $\hat{\boldsymbol\beta} \sim \mathcal{N}_q(\boldsymbol\beta\,;\,\sigma^2(\mathbf{X}'\mathbf{X})^{-1})$. Il est facile de montrer que

$$E(\hat{\boldsymbol\beta}'\hat{\boldsymbol\beta}) = \boldsymbol\beta'\boldsymbol\beta + \sigma^2 trace[(\mathbf{X}'\mathbf{X})^{-1}] = \boldsymbol\beta'\boldsymbol\beta + \sigma^2\sum_{i=1}^{q} 1/\lambda_i$$

où $\lambda_1, \lambda_2, \ldots, \lambda_q$ sont les valeurs propres (positives) de $\mathbf{X}'\mathbf{X}$. Si l'une d'elle est très petite, l'estimation $\hat{\boldsymbol\beta}$ sera de longueur exagérée. Si certaines valeurs propres de la matrice $\mathbf{X}'\mathbf{X}$ sont très petites comparées à d'autres, c'est que certaines variables sont fortement liées à d'autres. Diverses solutions sont proposées dans la littérature. Deux solutions possibles sont la régression en composantes principales et la régression *ridge*. Nous en dirons quelques mots plus tard.

### Moindres carrés pondérés

Il est possible de remplacer l'hypothèse $\mathrm{Var}(\boldsymbol\varepsilon) = \sigma^2\mathbf{I}_n$ par

$$\mathrm{Var}(\boldsymbol\varepsilon) = \sigma^2\mathbf{V},$$

lorsque $\mathbf{V}$ est une matrice définie positive *connue*.

Le critère des moindres carrés, selon lequel on minimise la somme des carrés $(\boldsymbol{y}-\mathbf{X}\boldsymbol\beta)'(\boldsymbol{y}-\mathbf{X}\boldsymbol\beta)$, doit être doit alors être modifié. La quantité qui doit être minimisée est plutôt

$$Q = (\boldsymbol{y}-\mathbf{X}\boldsymbol\beta)'\mathbf{V}^{-1}(\boldsymbol{y}-\mathbf{X}\boldsymbol\beta).$$

Étant définie positive, $\mathbf{V}$ possède une racine carrée symétrique $\mathbf{V}^{-½}$. Donc

$$Q = (\mathbf{V}^{-½}\boldsymbol{y}-\mathbf{V}^{-½}\mathbf{X}\boldsymbol\beta)'(\mathbf{V}^{-½}\boldsymbol{y}-\mathbf{V}^{-½}\mathbf{X}\boldsymbol\beta) = (\boldsymbol{z}-\mathbf{Y}\boldsymbol\beta)'(\boldsymbol{z}-\mathbf{Y}\boldsymbol\beta)$$

L'estimateur de $\boldsymbol\beta$ est

$$\hat{\boldsymbol\beta} = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\boldsymbol{z} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{y}$$

Cet estimateur est appelé l'estimateur des *moindres carrés pondéré.*
La somme des carrés résiduelle est

$$\mathrm{SCR} = (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol\beta})'\mathbf{V}^{-1}(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol\beta}) = \boldsymbol{y}'\left[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\boldsymbol{y}$$

L'estimateur de $\sigma^2$ est

$$\hat{\sigma}^2 = \frac{\mathrm{SCR}}{n-q} = \frac{1}{n-q}\boldsymbol{y}'\left[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right]\boldsymbol{y}$$

La statistique pour tester l'hypothèse $H_o : \mathbf{L}'\boldsymbol\beta = \mathbf{0}$ est

$$(\mathbf{L}'\hat{\boldsymbol\beta} - \mathbf{L}'\boldsymbol\beta)'[\mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L}]^{-1}(\mathbf{L}'\hat{\boldsymbol\beta} - \mathbf{L}'\boldsymbol\beta)/\hat{\sigma}^2$$

qui suit une loi $\mathcal{F}_{r,n-q}$ sous $H_o$.

Cette procédure équivaut à transformer le modèle $\boldsymbol{y} = \mathbf{X}\boldsymbol\beta + \boldsymbol\varepsilon$ en multipliant à gauche par $\mathbf{V}^{-½}$ :

$$\boldsymbol{y} = \mathbf{X}\boldsymbol\beta + \boldsymbol\varepsilon \Leftrightarrow \mathbf{V}^{-½}\boldsymbol{y} = \mathbf{V}^{-½}\mathbf{X}\boldsymbol\beta + \mathbf{V}^{-½}\boldsymbol\varepsilon \text{ ou } \boldsymbol{u} = \mathbf{Z}\boldsymbol\beta + \boldsymbol\delta$$

avec $\boldsymbol{u} = \mathbf{V}^{-½}\boldsymbol{y}$ ; $\mathbf{Z} = \mathbf{V}^{-½}\mathbf{X}$ ; et $\boldsymbol\delta = \mathbf{V}^{-½}\boldsymbol\varepsilon$. Le modèle transformé $\boldsymbol{u} = \mathbf{Z}\boldsymbol\beta + \boldsymbol\delta$ est alors le modèle classique traité ici puisque $\mathrm{Var}(\boldsymbol\delta) = \sigma^2\mathbf{I}_n$.

**Données centrées**

Il est parfois commode de « centrer » les données, c'est-à-dire, de soustraire de chacune la moyenne de la variable, car alors l'espérance des observations ne dépend plus de $\beta_o$, ce qui permet de se concentrer sur les autres composantes de $\boldsymbol{\beta}$. Ceci équivaut à multiplier les colonnes de données par une certaine matrice $\mathbf{C}$ définie par $\mathbf{C} = \mathbf{I}\text{-}\boldsymbol{ee}'/n$. Soit $\mathbf{X} = [\boldsymbol{e} \mid \mathbf{X}_1]$. On peut donc écrire le modèle sous la forme $\boldsymbol{y} = \beta_o\boldsymbol{e} + \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$. Sous forme centrée, le modèle devient $\boldsymbol{u} = \mathbf{Z}\boldsymbol{\beta}_1+\boldsymbol{\delta}$, où $\boldsymbol{u} = \mathbf{C}\boldsymbol{y}$, $\mathbf{Z} = \mathbf{C}\mathbf{X}_1$ et $\boldsymbol{\delta} = \mathbf{C}\boldsymbol{\varepsilon}$. Remarquez que $\boldsymbol{\delta}$ est de matrice de covariance $\sigma^2\mathbf{C}$, une matrice singulière : $\mathbf{C}$ est de rang $n$-1. Néanmoins, tout se déroule comme dans le modèle linéaire avec matrice de covariance scalaire : $\hat{\boldsymbol{\beta}}_1 = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{y} = (\mathbf{X}_1'\mathbf{C}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{C}\boldsymbol{y}$, de variance $\mathrm{Var}(\hat{\boldsymbol{\beta}}_1) = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} = \sigma^2(\mathbf{X}_1'\mathbf{C}\mathbf{X}_1)^{-1}$.

On développe ces propriétés en définissant une transformation équivalente mais non singulière. Puisque $\mathbf{C}$ est idempotente et symétrique elle peut s'écrire sous la forme $\mathbf{E}\mathbf{E}'$ où $\mathbf{E}$ est de dimensions $n\times(n\text{-}1)$ et $\mathbf{E}'\mathbf{E} = \mathbf{I}_{n\text{-}1}$ (les colonnes de $\mathbf{E}$ sont orthonormales). On définit la transformation $\mathbf{E}'\boldsymbol{y} = \mathbf{E}'\mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\eta}$ où $\boldsymbol{\eta} = \mathbf{E}'\boldsymbol{\varepsilon}$ est de moyenne nulle et de matrice de covariance $\sigma^2\mathbf{I}_{n\text{-}1}$. Ce modèle est alors le modèle classique avec matrice de covariance scalaire.