Postulats vérifiés:

$$(\mathbb{R}^p) \qquad \theta? \qquad Y = X\theta + \varepsilon$$

---

exemple: ANOVA 2 facteurs

$$\begin{array}{ccc} & A & \\ \swarrow & & \searrow \\ \alpha_1 & & \alpha_2 \end{array} \qquad \qquad \begin{array}{ccc} & B & \\ \swarrow & \downarrow & \searrow \\ \beta_1 & \beta_2 & \beta_3 \end{array}$$

Données:

		B: **1**	**2**	**3**
A	**1**	Y_{111} Y_{112}	Y_{121} Y_{122}	Y_{131} Y_{132}
	2	Y_{211} Y_{212}	Y_{221} Y_{222}	Y_{231} Y_{232}

modèle additif

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

On suppose les postulats vérifiés.
On cherche le vecteur θ, de $\mathbb{R}^p$, tel que $Y = X\theta + \varepsilon$.

Exemple d'une ANOVA à deux facteurs à deux et trois niveaux, avec deux répétitions.

$\Rightarrow$ 12 données: $Y_{111}, Y_{112}, \ldots, Y_{231}, Y_{232}$.

On choisit un modèle additif:

i: indice du facteur A,
j: indice du facteur B,
k: indice de repetition,
μ: effet moyen,
α_i: effet du niveau i de A,
β_j: effet du niveau j de B,
ε_{ijk}: erreurs.

$$
\begin{array}{ccccc}
Y & = & X\theta & + & \varepsilon
\end{array}
$$

$$
\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ Y_{232} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} . \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ \varepsilon_{232} \end{pmatrix}
$$

$$
\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ Y_{232} \end{pmatrix} = \mu \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta_2 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ \varepsilon_{232} \end{pmatrix}
$$

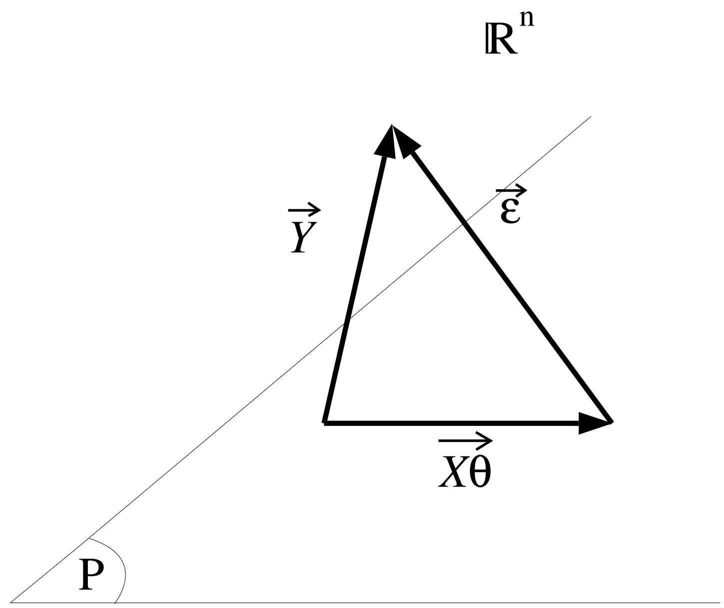
$$
\underset{\in \mathbb{R}^n}{\overrightarrow{Y}} = \underbrace{\mu.\ \overrightarrow{\mathbf{1}} + \alpha_1.\ \overrightarrow{A_1} + \alpha_2.\ \overrightarrow{A_2} + \beta_1.\ \overrightarrow{B_1} + \beta_2.\ \overrightarrow{B_2} + \beta_3.\ \overrightarrow{B_3}}_{\in \mathbb{R}^n} + \underset{\in \mathbb{R}^n}{\overrightarrow{\varepsilon}}
$$

- X: Matrice du plan d'expérience (n, p) avec $n = 12$ et $p = 6$.
- $\overrightarrow{Y}$: vecteur de $\mathbb{R}^{\mathrm{n}}$,
- $\vec{\varepsilon}$: vecteur de $\mathbb{R}^{\mathrm{n}}$,
- $\overrightarrow{X\theta}$: vecteur de $\mathbb{R}^{\mathrm{n}}$,
  combinaison linéaire des p vecteurs colonne de X:
  $\vec{1}, \overrightarrow{A_1}, \overrightarrow{A_2}, \overrightarrow{B_1}, \overrightarrow{B_2}, \overrightarrow{B_3}$.

Remarque: Par simplification d'écriture, on supprime parfois les flèches (exemple: $Y = X\theta + \varepsilon$).

$$\overrightarrow{Y} = \overrightarrow{X\theta} + \overrightarrow{\varepsilon}$$

$\overrightarrow{X\theta}$: combinaison linéaire des p vecteurs colonne de X.



P : sous-espace vectoriel de $\mathbb{R}^n$ engendré par les p vecteurs colonne de X.

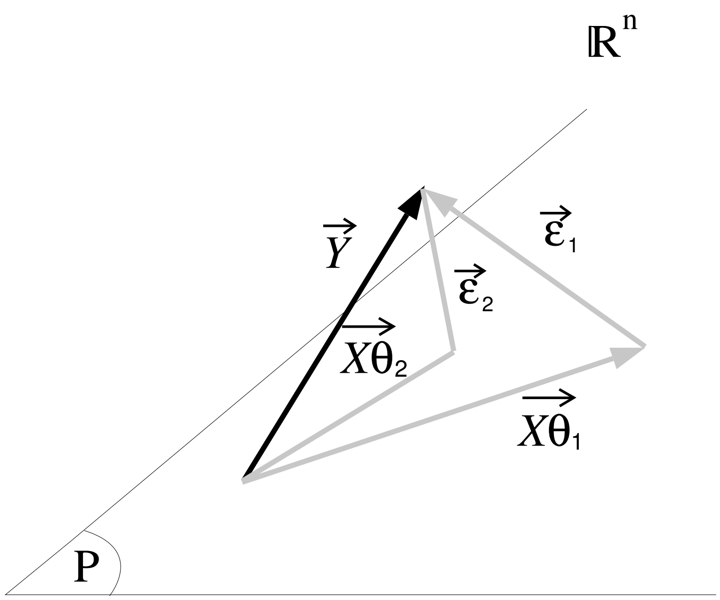
Soit P le sous-espace vectoriel de $\mathbb{R}^{\mathrm{n}}$,
engendré par les p vecteurs colonne de X.

Le vecteur $\overrightarrow{Y}$ est décomposé en:

- un vecteur $\overrightarrow{X\theta}$ de P,
- un vecteur $\overrightarrow{\varepsilon}$ de $\mathbb{R}^{\mathrm{n}}$.

$\overrightarrow{X\theta} \in P \subset \mathbb{R}^{\mathrm{n}}$

Voici une décomposition possible de $\overrightarrow{Y}$ en $\overrightarrow{X\theta}$ et $\overrightarrow{\varepsilon}$



Infinité de décompositions possibles.

Il existe une infinité de décompositions possibles.
Il nous faut donc un critère de choix:

On cherche donc **la** décomposition $\overrightarrow{Y} = \overrightarrow{X\hat{\theta}} + \overrightarrow{\hat{\varepsilon}}$ telle que:

- $\overrightarrow{X\hat{\theta}}$ ressemble le plus possible à $\overrightarrow{Y}$,
- on perde le moins possible d'informations en résumant $\overrightarrow{Y}$ par $\overrightarrow{X\hat{\theta}}$,
- les erreurs ε_i $(i = 1, \cdots, n)$ soient les plus petites possible(1).

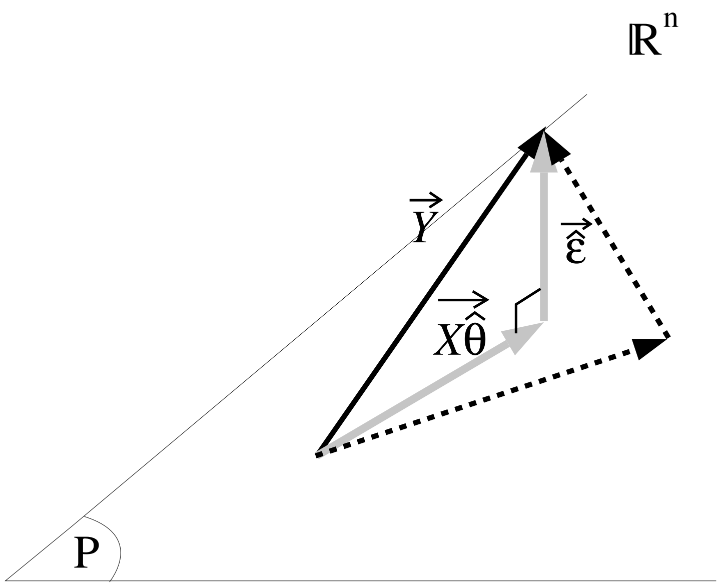
$\Rightarrow$ La longueur du vecteur $\overrightarrow{\varepsilon}$ soit la plus petite possible.

$\Rightarrow$ rendre $\| \overrightarrow{\varepsilon} \|^2$ minimale.

$$\boxed{\text{comment rendre } \| \overrightarrow{\varepsilon} \|^2 \text{ minimale?}}$$

(1) *Remarque formateur* : il y a ici une petie incohérence de notation : l'indice i de ε_i n'est pas le même que l'indice i de ε_{ijk}. Cette autre notation sera réutilisée par la suite : Y_i $(i = 1, \cdots, n)$ (exemple p.6).

$$\left\|\overrightarrow{\varepsilon}\right\|^2 \quad \text{minimale}$$

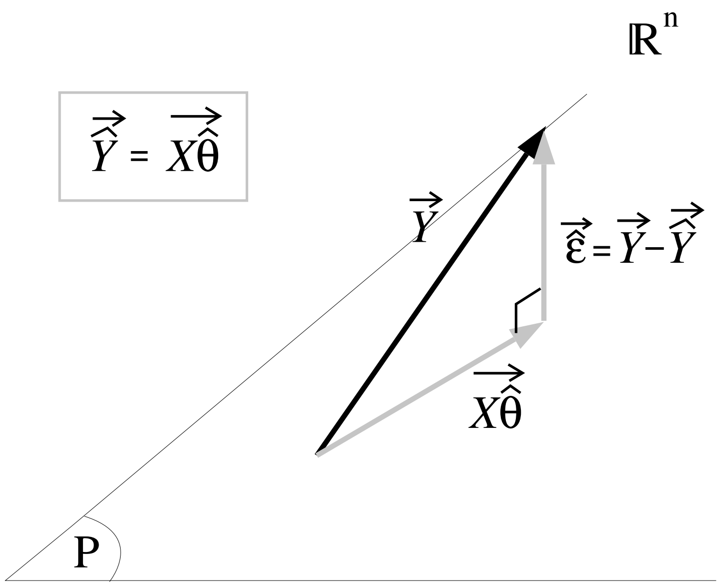


$$\Rightarrow \hat{\theta} \quad \text{tel que} \quad \overrightarrow{X\hat{\theta}} = \mathrm{Proj}^{\perp}_{\mathrm{P}}\left(\overrightarrow{Y}\right)$$

$\| \vec{\varepsilon} \|^2$ minimale lorsque $\overrightarrow{X\hat{\theta}} \perp \vec{\hat{\varepsilon}}$.

conclusion:

on obtiendra $\overrightarrow{X\hat{\theta}}$ en projetant orthogonalement $\overrightarrow{Y}$ sur P (sous-espace vectoriel engendré par les p vecteurs colonne de X).



$$\left\| \overrightarrow{\hat{\varepsilon}} \right\|^2 \quad \text{minimale}$$

$$\Leftrightarrow \left\| \overrightarrow{Y} - \overrightarrow{\widehat{Y}} \right\|^2 \quad \text{minimale}$$

$$\Leftrightarrow \sum_{i=1}^{n} \left(Y_i - \widehat{Y_i} \right)^2 \quad \text{minimale}$$

Projection $\perp$ = critère des moindres carrés

Projeter orthogonalement Y sur P,

- c'est rendre $\| \vec{\varepsilon} \|^2$ minimale,
- c'est donc rendre la somme des carrés des écarts entre valeurs observées et valeurs ajustées minimale,
- c'est minimiser le critère des moindres carrés.

conclusion:

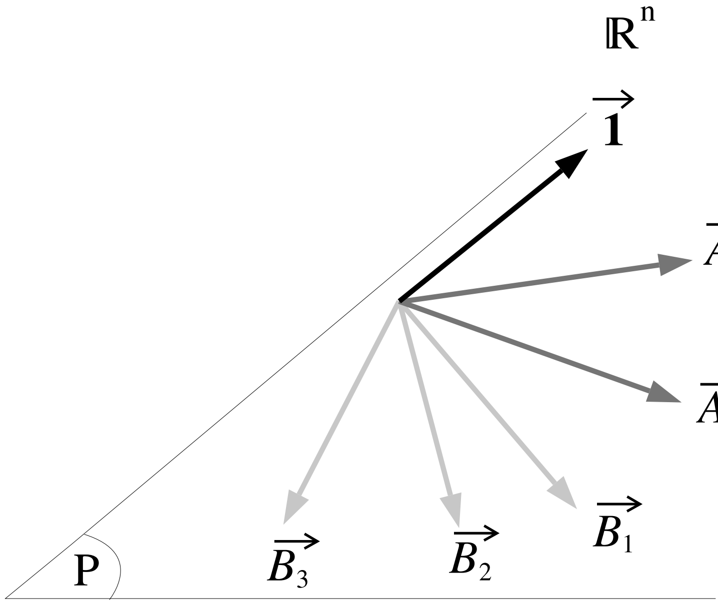
La meilleur estimation de $\overrightarrow{Y}$, au sens des moindres carrés, c'est la projection orthogonale de $\overrightarrow{Y}$ sur P,
où P est le sous-espace vectoriel de $\mathbb{R}^{\mathrm{n}}$, engendré par les p vecteurs colonne de X.

Rappel:

$\| \vec{a} \|^2 = < a, a >$
produit scalaire entre a et a.

$$\text{Or } < a, b > \; = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

Donc, $\|a\|^2 = \sum_{i=1}^{n} a_i^2$



P: engendré par les p vecteurs colonne de X:

$$\vec{\mathbf{1}}\ \overrightarrow{A_1},\ \overrightarrow{A_2},\ \overrightarrow{B_1},\ \overrightarrow{B_2},\ \overrightarrow{B_3}$$

Les p vecteurs colonne de X
$=$ base de P ?

remarque pour le formateur:
On va chercher les dimensions

- du sous-espace P,
- de $P^{\perp}$,
- de $\mathbb{R}^{\mathrm{n}}$,

et arriver à la notion de degré de liberté.

Rappel:
Des vecteurs $\vec{a_1} \ldots, \vec{a_n}$ forment une base $\mathcal{B}$ d'un espace vectoriel E, si tout vecteur $\vec{x}$ de E peut s'ecrire, de manière unique, sous forme d'une combinaison linéaire des vecteurs $\vec{a_1}, \ldots, \vec{a_n}$.
ou encore:

$\vec{a_1}, \ldots, \vec{a_n}$ forment une base si ces n vecteurs engendrent l'espace E, et si $\vec{a_1}, \ldots, \vec{a_n}$ sont linéairement indépendants (aucun des vecteurs $\vec{a_1}, \ldots, \vec{a_n}$ ne peut s'écrire comme combinaison linéaire des autres).

Les p vecteurs colonne de X forment-ils une base de P?
$\Longleftrightarrow$ Sont-ils linéairement indépendants?
$\Longleftrightarrow$ Existe-t-il des vecteurs colonne qui puissent s'exprimer comme une combinaison linéaire d'autres vecteurs colonne?

Les p vecteurs colonne de X = base de P ?

$$X = \begin{array}{c} \begin{array}{cccccc} \vec{\mathbf{1}} & \overrightarrow{A_1} & \overrightarrow{A_2} & \overrightarrow{B_1} & \overrightarrow{B_2} & \overrightarrow{B_3} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array} \quad (n,p)$$

$$\begin{cases} \overrightarrow{A_1} + \overrightarrow{A_2} = \vec{\mathbf{1}} \\ \overrightarrow{B_1} + \overrightarrow{B_2} + \overrightarrow{B_3} = \vec{\mathbf{1}} \end{cases} \quad \text{2 liaisons}$$

p vecteurs colonne

$p-2$ vecteurs colonne linéairement indépendants

$\Rightarrow$ base de P

$$\boxed{dim\ (\mathrm{P}) = p - 2}$$

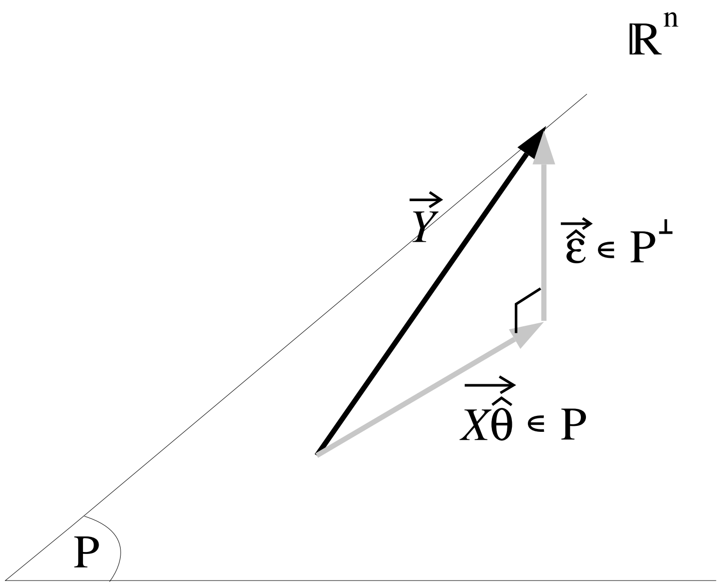
Existe-t-il des vecteurs colonne qui sont combinaisons linéaires d'autres vecteurs colonne?

Il existe deux relations liant les vecteurs colonne entre eux.
$\Longrightarrow$ Les p vecteurs colonne sont liés entre eux (par deux relations, car deux facteurs).

Donc, pour avoir une base de P, il faut $p-2$ vecteurs indépendants.(Il suffirait de supprimer une colonne par facteur, par exemple les colonnes $\overrightarrow{A_1}$ et $\overrightarrow{B_1}$).

Donc,

$$Dim(P) = p - 2 \qquad (= 4)$$
$$\overrightarrow{X\hat{\theta}} \in \mathbb{R}^{p-2}$$

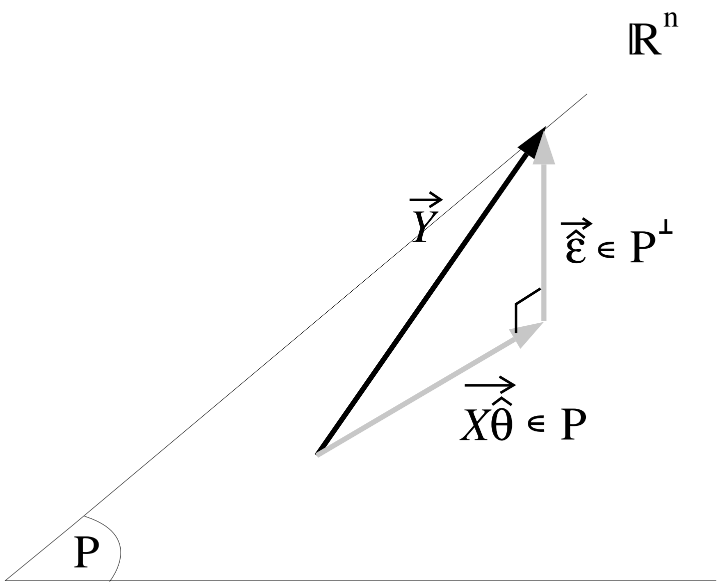


$$\mathbb{R}^n \quad = \quad \mathrm{P} \quad + \quad \mathrm{P}^{\perp}$$

$$\boxed{dim\ (\mathbb{R}^n)} \qquad \boxed{dim\ (\mathrm{P})} \qquad \boxed{dim\ (\mathrm{P}^{\perp})}$$

$$= n \qquad = p - 2 \qquad = ?$$

Version stagiaire incomplète



$$\mathbb{R}^n \quad = \quad \mathrm{P} \quad + \quad \mathrm{P}^{\perp}$$

$$\boxed{dim\ (\mathbb{R}^n)} = \boxed{dim\ (\mathrm{P})} + \boxed{dim\ (\mathrm{P}^{\perp})}$$

$$= n \qquad = p - 2 \qquad = n - (p-2)$$

On a décomposé $\mathbb{R}^n$ en deux sous-espaces vectoriels orthogonaux (d'intersection réduite au vecteur nul, c'est-à-dire que tout vecteur de l'un est perpendiculaire à tout vecteur de l'autre).

$$\mathbb{R}^n = P + P^{\perp}$$

$$dim\left(\mathbb{R}^n\right) = n$$
$$dim\left(P\right) = p - 2$$

quelle est la dimension de $P^{\perp}$?
$dim\left(P^{\perp}\right) = n - (p - 2)$

Pourquoi? Parce que:

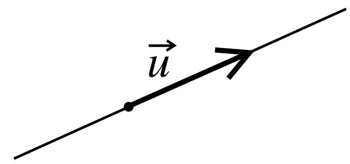
$$\boxed{dim\left(\mathbb{R}^n\right) = dim\left(P\right) + dim\left(P^{\perp}\right)}$$

et cela vient de l'orthogonalité de P et de $P^{\perp}$ (au sens de sous espaces orthogonaux d'intersection réduite au vecteur nul).

C'est donc vrai **par construction:**

Cela vient de la **projection orthogonale** de Y sur P.

Degré de liberté associé à $\overrightarrow{Y}$?



Un vecteur d'une droite : quel degré de liberté ? ...
Un vecteur d'un plan : quel degré de liberté ? ...
Un vecteur de $\mathbb{R}^n$: quel degré de liberté ? ...

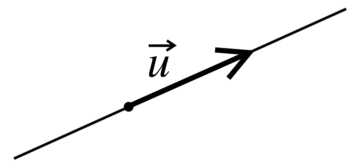
$\overrightarrow{Y}$: n coordonnées indépendantes
quel degré de liberté ? ...
d.l. de $\overrightarrow{Y}$ = ... = dim (...)

Degré de liberté = ...

$$\begin{array}{ccccc} \overrightarrow{Y} & = & \overrightarrow{X\hat{\theta}} & + & \overrightarrow{\hat{\varepsilon}} \\ \boxed{dim\,\mathbb{R}^n} & = & \boxed{dim\,(\mathrm{P})} & + & \boxed{dim\,(\mathrm{P}^{\perp})} \\ \boxed{\text{d.l. de } \overrightarrow{Y}} & & \boxed{\text{d.l. de } \overrightarrow{X\hat{\theta}}} & & \boxed{\text{d.l. de } \overrightarrow{\hat{\varepsilon}}} \\ = \mathrm{n} & & = \ldots & & = \ldots \end{array}$$

Version stagiaire incomplète

$\boxed{\textbf{Degré de liberté associé à } \overrightarrow{Y} \textbf{ ?}}$



Un vecteur d'une droite : quel degré de liberté ? 1
Un vecteur d'un plan : quel degré de liberté ? 2
Un vecteur de $\mathbb{R}^n$: quel degré de liberté ? n

$\overrightarrow{Y}$: n coordonnées indépendantes
quel degré de liberté ? n

$$\text{d.l. de } \overrightarrow{Y} = \text{n} = dim\ (\ \mathbb{R}^n\ )$$

Degré de liberté =	**dimension d'E.V.**

$$\begin{array}{ccccc}
\overrightarrow{Y} & = & \overrightarrow{X\hat{\theta}} & + & \overrightarrow{\hat{\varepsilon}} \\
\boxed{dim\,\mathbb{R}^n} & = & \boxed{dim\ (\mathrm{P})} & + & \boxed{dim\ (\mathrm{P}^{\perp})} \\
\boxed{\text{d.l. de } \overrightarrow{Y}} & = & \boxed{\text{d.l. de } \overrightarrow{X\hat{\theta}}} & + & \boxed{\text{d.l. de } \overrightarrow{\hat{\varepsilon}}} \\
= \text{n} & & = p-2 & & = n-(p-2)
\end{array}$$

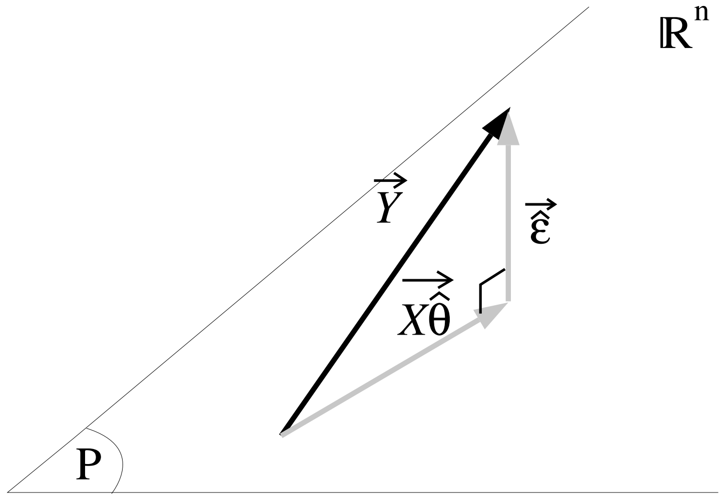
Quel est le degré de liberté associé au vecteur $\vec{Y}$?

Rappels:
Un vecteur ne pouvant se déplacer que sur une droite, n'a qu'un degré de liberté: une seule coordonnée peut varier.
Un vecteur d'un plan a deux degrés de liberté, deux coordonnées peuvent varier...

$\vec{Y}$ a n coordonnées indépendantes car il y a n mesures indépendentes, (d'où le postulat d'indépendance des ε_i).
$\vec{Y}$ a donc n degrés de liberté pour se mouvoir dans $\mathbb{R}^n$.

Le degré de liberté associé à $\vec{Y}$, c'est la dimension de l'espace vectoriel dans lequel $\vec{Y}$ peut se mouvoir (ici $\mathbb{R}^n$ tout entier).

Les degrés de liberté sont des dimensions de sous-espaces vectoriels. Comme les dimensions des sous-espaces vectoriels P et $P^{\perp}$ se somment pour donner la dimension de $\mathbb{R}^n$, il en est de même pour les degrés de liberté. (et cela est vrai par construction: on obtient $\vec{\hat{Y}}$ par projection orthogonale de $\vec{Y}$ sur P).



$\overrightarrow{Y}$	$=$	$\overrightarrow{X\hat{\theta}}$	$+$	$\overrightarrow{\hat{\varepsilon}}$
$\in$		$\in$		$\in$
$\mathbb{R}^n$	$=$	P	$+$	$\mathrm{P}^{\perp}$
$\boxed{dim\ (\mathbb{R}^n)}$	$=$	$\boxed{dim\ (\mathrm{P})}$	$+$	$\boxed{dim\ (\mathrm{P}^{\perp})}$
$\boxed{\text{d.l. de } \overrightarrow{Y}}$	$=$	$\boxed{\text{d.l. de } \overrightarrow{X\hat{\theta}}}$	$+$	$\boxed{\text{d.l. de } \overrightarrow{\hat{\varepsilon}}}$
$\boxed{\left\lVert\overrightarrow{Y}\right\rVert^2}$	$=$	$\boxed{\left\lVert\overrightarrow{X\hat{\theta}}\right\rVert^2}$	$+$	$\boxed{\left\lVert\overrightarrow{\hat{\varepsilon}}\right\rVert^2}$

car Pythagore
projection orthogonale de $\overrightarrow{Y}$ sur P.

Cela vient du théorème de Pythagore:
en effet, $\vec{\hat{\varepsilon}} \perp \overrightarrow{X\hat{\theta}}$ par construction (projection orthogonale de $\overrightarrow{Y}$ sur P).

Projection orthogonale de $\overrightarrow{Y}$ sur P

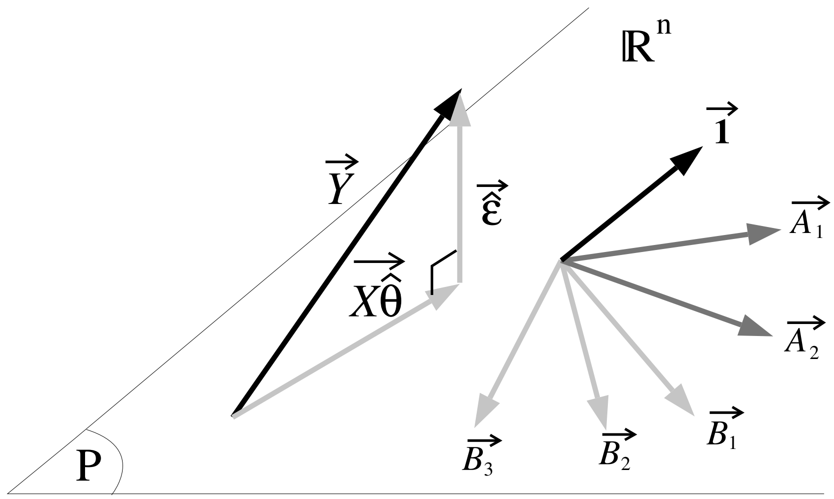


Table d'analyse de la variance:

Source de variation	Degrés de liberté	Sommes des carrés
(expliquée par le) modèle	$dim\ (\mathrm{P}) = p - 2$	$\left\Vert \overrightarrow{X\hat{\theta}} \right\Vert^2$
Résiduelle	$dim\ (\mathrm{P}^{\perp}) = n-(p-2)$	$\left\Vert \overrightarrow{\hat{\varepsilon}} \right\Vert^2$
Totale	$dim\ (\mathbb{R}^n) = n$	$\left\Vert \overrightarrow{Y} \right\Vert^2$

Voila ce qu'on a fait:

- On a projeté $\overrightarrow{Y}$ sur le sous-espace vectoriel P engendré par les vecteurs colonnes de X (projection orthogonale).

- on peut résumer cela dans un tableau d'analyse de la variance.

- On vérifie dans ce tableau, que les degrés de liberté "modèle" et "résiduelle" se somment pour donner le degré de liberté "total". De même pour les sommes de carrés.

- et ceci vient de la construction de $\overrightarrow{\hat{Y}}$ = projection orthogonale de $\overrightarrow{Y}$ sur P.

Projection orthogonale de $\overrightarrow{Y} - \overline{Y}.\, \overrightarrow{\mathbf{1}}$ sur P'

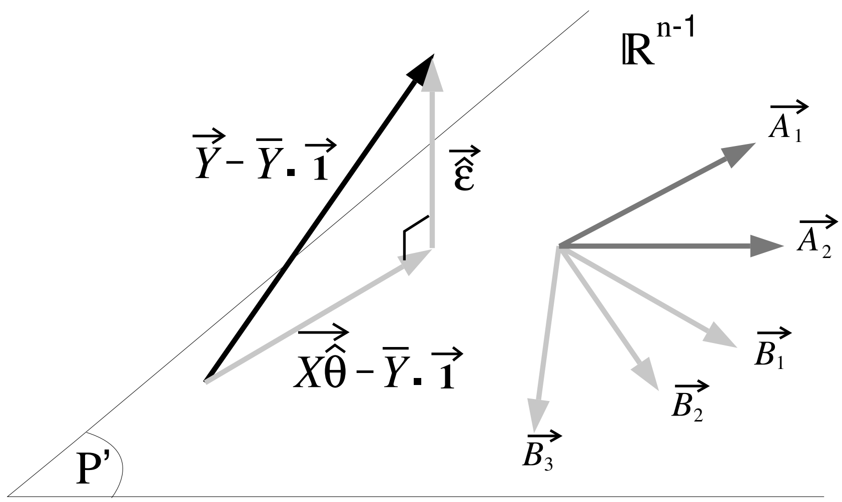


Table d'analyse de la variance:

Source de variation	Degrés de liberté	Sommes des carrés
(expliquée par le) modèle	$dim\ (\mathrm{P}) - 1 = p - 3$	$\left\Vert \overrightarrow{X\hat{\theta}} - \overline{Y}.\, \overrightarrow{\mathbf{1}} \right\Vert^2$
Résiduelle	$dim\ \left(\mathrm{P}^{\perp}\right) = n - 1 - (p - 3)$	$\left\Vert \overrightarrow{\hat{\varepsilon}} \right\Vert^2$
Totale	$dim\ \left(\mathbb{R}^{n-1}\right) = n - 1$	$\left\Vert \overrightarrow{Y} - \overline{Y}.\, \overrightarrow{\mathbf{1}} \right\Vert^2$

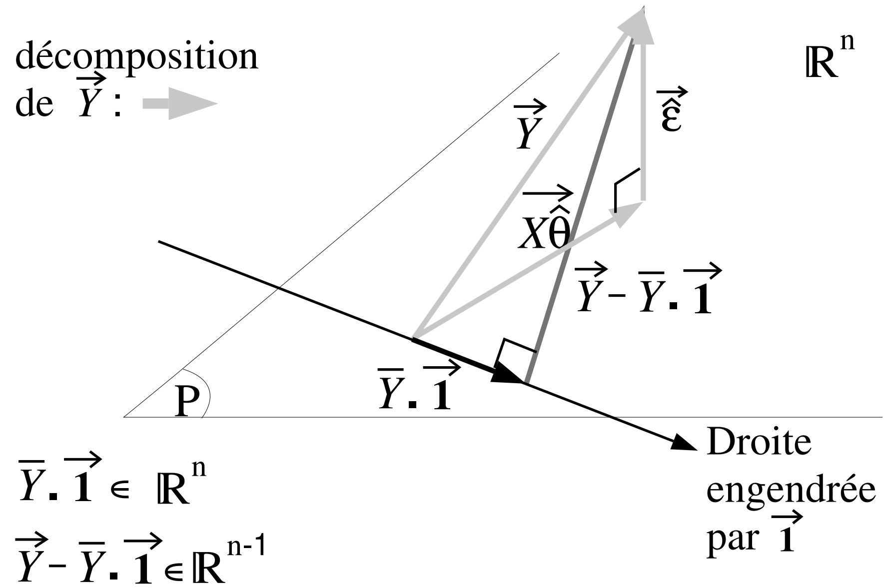
En fait, on ne fait pas tout à fait comme ça.

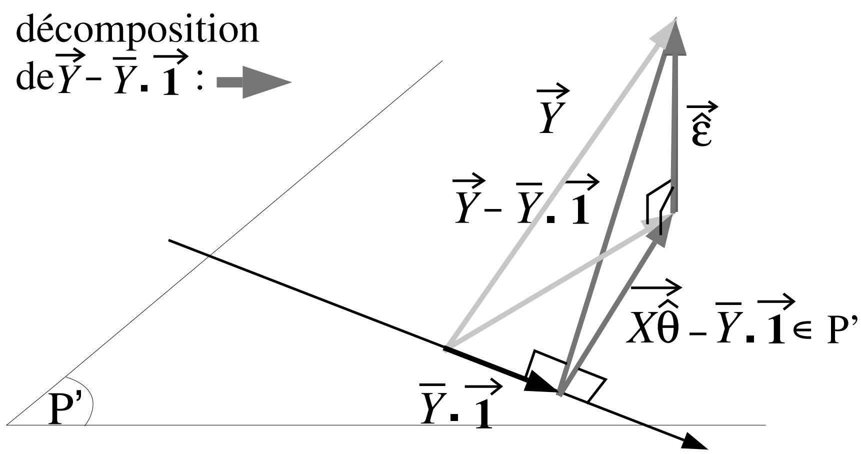
$\overline{Y}$ (la moyenne générale) n'apporte aucune explication au modèle, au phénomène.

- On va donc projeter $\overrightarrow{Y} - \overline{Y}.\,\vec{\mathbf{1}}$ orthogonalement, sur le sous espace vectoriel engendré par les vecteurs $\overrightarrow{A_1}, \overrightarrow{A_2}, \overrightarrow{B_1}, \overrightarrow{B_2}, \overrightarrow{B_3}$.(de dimension égale à $dim(P) - 1$).

- On obtient la nouvelle table d'analyse de variance:
  - avec un degré de liberté de moins pour les sommes de carrés "modèle" et "totale",
  - Les degrés de liberté et les sommes de carrés se somment toujours: les sous-espaces vectoriels sur lesquels on décompose $\overrightarrow{Y} - \overline{Y}.\,\vec{\mathbf{1}}$ sont toujours orthogonaux.

Décomposition de $\overrightarrow{Y}$ et $\overrightarrow{Y} - \overline{Y}.\overrightarrow{\mathbf{1}}$





En gris clair: la décomposition de $\overrightarrow{Y}$ en un vecteur $\overrightarrow{X\hat{\theta}}$ de P et un vecteur $\vec{\hat{\varepsilon}}$ de $P^{\perp}$.

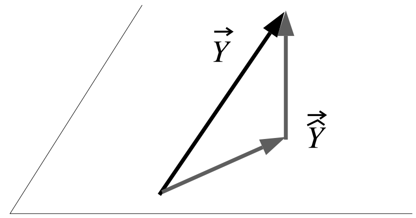
- On représente, dans P, la droite engendrée par $\vec{\mathbf{1}}$, le vecteur $\overline{Y}_{\bullet}\,\vec{\mathbf{1}}$ sur cette droite (= projection orthogonale de $\overrightarrow{Y}$ sur la droite).
- On représente ensuite le vecteur $\overrightarrow{Y} - \overline{Y}_{\bullet}\,\vec{\mathbf{1}}$, que l'on veut décomposer en un vecteur de $P' \subset P$, engendré par $\overrightarrow{A_1}, \overrightarrow{A_2}$ $, \overrightarrow{B_1}, \overrightarrow{B_2}, \overrightarrow{B_3}$ et un vecteur orthogonal.

En gris foncé : la décomposition de $\overrightarrow{Y} - \overline{Y}_{\bullet}\,\vec{\mathbf{1}}$.
On obtient le vecteur $\overrightarrow{X\hat{\theta}} - \overline{Y}_{\bullet}\,\vec{\mathbf{1}}$ de $P' \subset P$, et le même vecteur $\vec{\hat{\varepsilon}}$ que précédemment.

Habituellement, c'est la deuxième décomposition que l'on trouve dans les tables d'analyse de variance. Dans un cas, on projette $\overrightarrow{Y}$, dans l'autre, on projette $\overrightarrow{Y} - \overline{Y}_{\bullet}\,\vec{\mathbf{1}}$. Généralement, on simplifie en disant qu'on projette $\overrightarrow{Y}$. par la suite on dira qu'on projette $\overrightarrow{Y}$ sur P, sachant bien que la table d'analyse de variance présente en fait la projection de $\overrightarrow{Y} - \overline{Y}_{\bullet}\,\vec{\mathbf{1}}$.

Exprimer $\hat{\theta}$ en fonction de Y

$$\widehat{Y} = X\hat{\theta}$$



$\hat{\theta}$?

$$\cancel{\hat{\theta} = X^{-1}\widehat{Y}}$$

$$X(n,p)$$

$$X'\widehat{Y} = X'X\hat{\theta} \qquad X'X(p,p)$$

or $\quad X'\hat{\varepsilon} = \overrightarrow{0}$

Donc, $\quad X'\left(\widehat{Y} + \hat{\varepsilon}\right) = X'X\hat{\theta}$

Donc, $\quad X'Y = X'X\hat{\theta}$

D'où $\quad \boxed{\hat{\theta} = (X'X)^{-1}X'Y}$

si $X'X$ inversible

- On a vu comment décomposer $\overrightarrow{Y}$ en $\overrightarrow{X\hat{\theta}}$ et $\vec{\hat{\varepsilon}}$, ou $\overrightarrow{Y} - \overline{Y}.\,\vec{\mathbf{1}}$ en $\overrightarrow{X\hat{\theta}} - \overline{Y}.\,\vec{\mathbf{1}}$ et $\vec{\hat{\varepsilon}}$.
- On a vu les tables d'analyse de variance associées.
- Mais on n'a toujours pas exprimé $\hat{\theta}$ en fonction du vecteur $\overrightarrow{Y}$ des observations.

on a $\overrightarrow{\hat{Y}} = X\ \overrightarrow{\hat{\theta}}$.
On supprimera maintenant les flèches au dessus des vecteurs comme on le fait traditionnellement.
D'où: $\hat{Y} = X\hat{\theta}$.

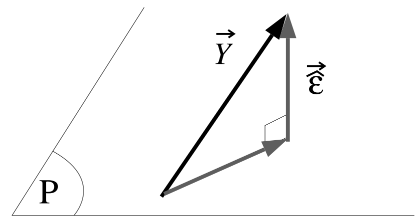
Peut-on écrire $\hat{\theta} = X^{-1}\hat{Y}$?
NON, $X(n,p)$ n'est pas carrée.
on peut écrire $\qquad X'\hat{Y} = X'X\hat{\theta}$.
or, $\qquad X'\hat{\varepsilon} = \vec{0} \qquad (1)$
(car $\hat{\varepsilon} \in P^{\perp}$, voir transparent suivant).
Donc, $X'(\hat{Y} + \hat{\varepsilon}) = X'X\hat{\theta}$.
Soit $X'Y = X'X\hat{\theta}$

$X'X(p,p)$: carrée.
Si $X'X$ est inversible, $(X'X)^{-1}$ existe, et on peut écrire:

$$\hat{\theta} = (X'X)^{-1}X'Y$$

(On laisse de côté le problème de l'inversibilité de $X'X$ pour le moment).

$$\overrightarrow{X'}\hat{\varepsilon} = \overrightarrow{0}$$



Démonstration :

$$\overrightarrow{\hat{\varepsilon}} \perp \overrightarrow{X_i} \quad \forall i = 1 \cdots p$$

$\overrightarrow{X_i}$: vecteurs colonnes de X engendrant P

$$< \overrightarrow{X_i}, \overrightarrow{\hat{\varepsilon}} > = 0 \quad \forall i = 1 \cdots p$$

$$\overrightarrow{X_i'} . \vec{\hat{\varepsilon}} = 0 \qquad X_i' \begin{pmatrix} . & . & . & . & . \end{pmatrix} \quad \vec{\hat{\varepsilon}} \begin{pmatrix} . \\ . \\ . \\ . \\ . \end{pmatrix} = 0 \qquad \forall \mathrm{i}$$

$$\text{Donc,} \quad X' \vec{\hat{\varepsilon}} = \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} \quad \begin{pmatrix} . \\ . \\ . \\ . \\ . \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \vec{0}$$

TRANSPARENT FACULTATIF: démonstration de $\overrightarrow{X'\hat{\varepsilon}}=\overrightarrow{0}$

On a utilisé (1): $\overrightarrow{X'\hat{\varepsilon}}=\overrightarrow{0}$, en disant que cela venait du fait que $\overrightarrow{\hat{\varepsilon}} \in P^{\perp}$.
Montrons le:

$$\overrightarrow{\hat{\varepsilon}} \in P^{\perp}$$

Donc, $\overrightarrow{\hat{\varepsilon}}$ est orthogonal à tout vecteur de P, en particulier aux vecteurs colonnes de X, que l'on note $\overrightarrow{X_i}$.
Donc, $\overrightarrow{\hat{\varepsilon}} \perp \overrightarrow{X_i} \qquad \forall i = 1 \dots p$
ce qui s'écrit $<\overrightarrow{X_i}\, ,\, \overrightarrow{\hat{\varepsilon}}>= 0 \qquad \forall i$
(Les produits scalaires entre les vecteurs $\overrightarrow{X_i}$ et $\overrightarrow{\hat{\varepsilon}}$ sont tous nuls).
cela s'écrit encore $\overrightarrow{X_i'}\ .\ \overrightarrow{\hat{\varepsilon}} = 0$ par définition du produit scalaire.
Soit

$$X_i' \begin{pmatrix} . & . & . & . & . \end{pmatrix} \quad \vec{\hat{\varepsilon}} \begin{pmatrix} . \\ . \\ . \\ . \\ . \end{pmatrix} = 0 \qquad \forall \text{i}$$

$$\text{Donc,} \quad X'\vec{\hat{\varepsilon}} = \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} \quad \begin{pmatrix} . \\ . \\ . \\ . \\ . \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \vec{0}$$

$$
\begin{aligned}
X'X \text{ inversible } &\text{ssi } X'X \text{ de plein rang}\\
&\text{ssi } X \text{ de plein rang}
\end{aligned}
$$

$$X \longrightarrow \widetilde{X} \text{ de plein rang}$$

$$\boxed{\hat{\theta} = \left(\widetilde{X}'\widetilde{X}\right)^{-1} \widetilde{X}'Y}$$

On a vu qu'on peut écrire

$$\hat{\theta} = (X'X)^{-1}X'Y$$

donc exprimer $\hat{\theta}$ en fonction de Y, si $X'X$ est inversible.

Dire que $X'X$ est inversible équivaut à dire que $X'X$ est de plein rang, c'est-à-dire qu'aucun vecteur ligne de $X'X$ n'est combinaison linéaire d'autres vecteurs lignes de $X'X$, de même qu'aucun vecteur colonne de $X'X$ n'est combinaison linéaire d'autres vecteurs colonne de $X'X$.

$$\text{or,} \qquad rang(X'X) = rang(X).$$

Donc $X'X$ est de plein rang si et seulement si X est de plein rang.

Nous avons vu (en ANOVA) que X n'est pas de plein rang puisque:

$$\overrightarrow{A_1} + \overrightarrow{A_2} = \overrightarrow{\mathbf{1}} \qquad \text{et} \qquad \overrightarrow{B_1} + \overrightarrow{B_2} + \overrightarrow{B_3} = \overrightarrow{\mathbf{1}}$$

L'ordinateur rend X de plein rang, en supprimant éventuellement des vecteurs colonne **qui créent des colinéarités**.
on remplace X par $\tilde{X}$, matrice de plein rang.

$$X \longrightarrow \widetilde{X} \text{ de plein rang}$$

$$\boxed{\hat{\theta} = \left(\widetilde{X}'\widetilde{X}\right)^{-1} \widetilde{X}'Y}$$

$$\widetilde{X}\,\hat{\theta} = \underbrace{\widetilde{X}\ \left(\widetilde{X}'\widetilde{X}\right)^{-1} \widetilde{X}'}_{\text{projecteur orthogonal sur P}}\ Y = \widehat{Y}$$

La matrice$X$ est remplacée par la matrice $\tilde{X}$ de plein rang.

On appelle $\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ le projecteur orthogonal sur P.

En effet,la matrice $\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ projette orthogonalement le vecteur Y sur P.

On peut démontrer que $\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ est un projecteur orthogonal:
c'est un projecteur orthogonal si et seulement si

- c'est un projecteur
- $\hat{Y} \perp \hat{\varepsilon}$.

$\boxed{\text{Rendre } X \text{ de plein rang}}$

Première solution :

$$
\begin{array}{ccccc}
Y & = & X\theta & + & \varepsilon \\
\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ Y_{232} \end{pmatrix} & = & \begin{array}{c} \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \\ \overrightarrow{1} \; \overrightarrow{A_1} \; \overrightarrow{A_2} \; \overrightarrow{B_1} \; \overrightarrow{B_2} \; \overrightarrow{B_3} \end{array} \cdot \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} & + & \begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ \varepsilon_{232} \end{pmatrix}
\end{array}
$$

$$
\begin{cases} \overrightarrow{A_1} + \overrightarrow{A_2} = \overrightarrow{\mathbf{1}} \\ \overrightarrow{B_1} + \overrightarrow{B_2} + \overrightarrow{B_3} = \overrightarrow{\mathbf{1}} \end{cases}
$$

Supprimer $\overrightarrow{A_1}$ et $\overrightarrow{B_1} \Longrightarrow \widetilde{X}$ de plein rang

$\vec{\mathbf{1}}$ est combinaison linéaire de $\overrightarrow{A_1}$ et $\overrightarrow{A_2}$.

$\vec{\mathbf{1}}$ est combinaison linéaire de $\overrightarrow{B_1}$, $\overrightarrow{B_2}$ et $\overrightarrow{B_3}$.

X n'est pas de plein rang.

Pour la rendre de plein rang il suffit de supprimer les colinéarités, et donc de supprimer un vecteur colonne du bloc A ($\overrightarrow{A_1}$ ou $\overrightarrow{A2}$), et l'un des vecteurs colonne du bloc B ($\overrightarrow{B_1}$ ou $\overrightarrow{B_2}$ ou $\vec{B_3}$).

On choisit arbitrairement de supprimer $\overrightarrow{A_1}$ et $\vec{B_1}$.

$\boxed{\text{Rendre } X \text{ de plein rang}}$

$$
\begin{array}{cccccc}
\multicolumn{6}{c}{\begin{pmatrix}
1 & . & 0 & . & 0 & 0 \\
1 & . & 0 & . & 0 & 0 \\
1 & . & 0 & . & 1 & 0 \\
1 & . & 0 & . & 1 & 0 \\
1 & . & 0 & . & 0 & 1 \\
1 & . & 0 & . & 0 & 1 \\
1 & . & 1 & . & 0 & 0 \\
1 & . & 1 & . & 0 & 0 \\
1 & . & 1 & . & 1 & 0 \\
1 & . & 1 & . & 1 & 0 \\
1 & . & 1 & . & 0 & 1 \\
1 & . & 1 & . & 0 & 1
\end{pmatrix}} \\
\overrightarrow{\mathbf{1}} & \overrightarrow{A_1} & \overrightarrow{A_2} & \overrightarrow{B_1} & \overrightarrow{B_2} & \overrightarrow{B_3}
\end{array}
\cdot
\begin{pmatrix}
\mu \\ \cancel{\alpha_1} \\ \alpha_2 \\ \cancel{\beta_1} \\ \beta_2 \\ \beta_3
\end{pmatrix}
$$

$$X \longrightarrow \widetilde{X}$$

$$\theta \longrightarrow \theta' = \begin{pmatrix} \mu' \\ \alpha_2' \\ \beta_2' \\ \beta_3' \end{pmatrix}$$

$X \longrightarrow \widetilde{X}$ $\iff$ fixer la paramétrisation

$\iff$ ajouter les contraintes $\begin{cases} \alpha_1 & = 0 \\ \beta_1 & = 0 \end{cases}$

Transformer X en $\tilde{X}$ revient à:

- transformer θ en θ',

- fixer la paramétrisation en ajoutant **les contraintes**:

$$\begin{cases} \alpha_1 = 0 \\ \beta_1 = 0 \end{cases}$$

$\boxed{\text{Rendre } X \text{ de plein rang}}$

$$X\theta = \widetilde{X}\,\theta'$$

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 1 & . & 0 & . & 0 & 0 \\ 1 & . & 0 & . & 0 & 0 \\ 1 & . & 0 & . & 1 & 0 \\ 1 & . & 0 & . & 1 & 0 \\ 1 & . & 0 & . & 0 & 1 \\ 1 & . & 0 & . & 0 & 1 \\ 1 & . & 1 & . & 0 & 0 \\ 1 & . & 1 & . & 0 & 0 \\ 1 & . & 1 & . & 1 & 0 \\ 1 & . & 1 & . & 1 & 0 \\ 1 & . & 1 & . & 0 & 1 \\ 1 & . & 1 & . & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mu' \\ . \\ \alpha_2' \\ . \\ \beta_2' \\ \beta_3' \end{pmatrix}$$

$$\begin{cases} \mu' & = \mu + \alpha_1 + \beta_1 \\ \alpha_2' & = \alpha_2 - \alpha_1 \\ \beta_2' & = \beta_2 - \beta_1 \\ \beta_3' & = \beta_3 - \beta_1 \end{cases} \Longrightarrow$$

passer d'une paramétrisation à une autre

TRANSPARENT FACULTATIF

On transforme X en $\tilde{X}$ et θ en θ'.
Mais on ne modifie pas les espérances.
Donc on a:

$$X\theta = \tilde{X}\theta'$$

D'où le système d'équations qui nous permet de passer d'une paramétrisation générale, à la paramétrisation fixée en choisissant les contraintes

$$\begin{cases} \alpha_1 = 0 \\ \beta_1 = 0 \end{cases}$$

$\boxed{\text{Rendre } X \text{ de plein rang}}$

Deuxième solution :

Contraintes $\begin{cases} \alpha_1 + \alpha_2 & = 0 \\ \beta_1 + \beta_2 + \beta_3 & = 0 \end{cases}$

$$\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ Y_{232} \end{pmatrix} = \begin{array}{c} \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \\ \begin{array}{cccccc} \overrightarrow{1} & \overrightarrow{A_1} & \overrightarrow{A_2} & \overrightarrow{B_1} & \overrightarrow{B_2} & \overrightarrow{B_3} \end{array} \end{array} . \begin{pmatrix} \mu \\ -\alpha_2 \\ \alpha_2 \\ -(\beta_2 + \beta_3) \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ \varepsilon_{232} \end{pmatrix}$$

$$\overrightarrow{X\theta} = \mu.\ \overrightarrow{\mathbf{1}}\ -\alpha_2.\ \overrightarrow{A_1}\ +\alpha_2.\ \overrightarrow{A_2} - (\beta_2 + \beta_3).\ \overrightarrow{B_1}\ +\beta_2.\ \overrightarrow{B_2}\ +\beta_3.\ \overrightarrow{B_3}$$

$$\overrightarrow{X\theta} = \mu.\ \overrightarrow{\mathbf{1}}\ +\alpha_2.(\overrightarrow{A_2} - \overrightarrow{A_1}) + \beta_2.(\overrightarrow{B_2} - \overrightarrow{B_1}) + \beta_3.(\overrightarrow{B_3} - \overrightarrow{B_1})$$

$$\begin{cases} \overrightarrow{\mathbf{1}} & \longrightarrow \overrightarrow{\mathbf{1}} \\ \overrightarrow{A_2} & \longrightarrow \overrightarrow{A_2} - \overrightarrow{A_1} \\ \overrightarrow{B_2} & \longrightarrow \overrightarrow{B_2} - \overrightarrow{B_1} \\ \overrightarrow{B_3} & \longrightarrow \overrightarrow{B_3} - \overrightarrow{B_1} \end{cases}$$

Autre solution:
Utiliser un autre système de contraintes, plus naturel:

$$\begin{cases} \alpha_1 + \alpha_2 = 0 \\ \beta_1 + \beta_2 + \beta_3 = 0 \end{cases}$$

On introduit les contraintes dans $\overrightarrow{\theta}$.

Utiliser ces contraintes revient à transformer X en $\tilde{X}$ en remplaçant les colonnes:

$$\begin{array}{lll} \overrightarrow{\mathbf{1}} & \text{par} & \overrightarrow{\mathbf{1}} \\ \overrightarrow{A_2} & \text{par} & \overrightarrow{A_2} - \overrightarrow{A_1} \\ \overrightarrow{B_2} & \text{par} & \overrightarrow{B_2} - \overrightarrow{B_1} \\ \overrightarrow{B_3} & \text{par} & \overrightarrow{B_3} - \overrightarrow{B_1} \end{array}$$

et en supprimant les colonnes $\overrightarrow{A_1}$ et $\overrightarrow{B_1}$.

Cela s'appelle recentrer la matrice X.

$\boxed{\text{Rendre } X \text{ de plein rang}}$

$$\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ Y_{232} \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 & 0 \\ 1 & -1 & -1 & 0 \\ 1 & -1 & 0 & -1 \\ 1 & -1 & 0 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ \varepsilon_{232} \end{pmatrix}$$

$\downarrow$

$\mathbf{1}$

$\downarrow$

$A_2 - A_1$

$B_2 - B_1 \quad B_3 - B_1$

Les vecteurs sont orthogonaux d'un bloc ($\mathbf{1}$, A ou B) à l'autre

$$\widetilde{X} = X \text{ recentrée}$$

$$X \longrightarrow \widetilde{X} \iff \text{fixer la paramétrisation}$$

$$\iff \text{imposer les contraintes} \begin{cases} \alpha_1 + \alpha_2 & = 0 \\ \beta_1 + \beta_2 + \beta_3 & = 0 \end{cases}$$

$\tilde{X}$ est de plein rang.

La matrice $\tilde{X}$ est composée de trois blocs:

- $\overrightarrow{\mathbf{1}}$
- $\overrightarrow{A_2} - \overrightarrow{A_1}$
- $\overrightarrow{B_2} - \overrightarrow{B_1}, \quad \overrightarrow{B_3} - \overrightarrow{B_1}$

Tout vecteur d'un bloc est orthogonal à tout vecteur d'un autre bloc:
On a décomposé P en trois sous-espaces vectoriels orthogonaux associés:

- à $\overrightarrow{\mathbf{1}}$,
- au facteur A,
- au facteur B.

$\boxed{\text{Rendre } X \text{ de plein rang}}$

$$
\overset{\widetilde{X}'}{\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & -1 & 0 & 0 & 1 & 1 & -1 & -1 & 0 & 0 & 1 & 1 \\
0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 & -1 & -1 & 1 & 1
\end{pmatrix}}
\;.\;
\overset{\widetilde{X}}{\begin{pmatrix}
1 & -1 & -1 & 0 \\
1 & -1 & -1 & 0 \\
1 & -1 & 0 & -1 \\
1 & -1 & 0 & -1 \\
1 & -1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 0 \\
1 & 1 & -1 & 0 \\
1 & 1 & 0 & -1 \\
1 & 1 & 0 & -1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1
\end{pmatrix}}
$$

$$
\widetilde{X}'\widetilde{X} = \left|\begin{array}{c|c|cc}
12 & 0 & 0 & 0 \\ \hline
0 & 12 & 0 & 0 \\ \hline
0 & 0 & 8 & 4 \\
0 & 0 & 4 & 8
\end{array}\right|
$$

bloc diagonale , plan orthogonal

1 bloc pour chaque terme du modèle :

$$
Y_{ijk} = \underset{\downarrow}{\mu} + \underset{\downarrow}{\alpha_i} + \underset{\downarrow}{\beta_j} + \varepsilon_{ijk}
$$

3 blocs = 3 s.e.v. orthogonaux

TRANSPARENT FACULTATIF

La matrice $\tilde{X}'\tilde{X}$ est bloc-diagonale (ou diagonale par bloc).

P est décomposé en trois sous-espaces orthogonaux:
un sous-espace vectoriel par terme du modèle:

$$
\begin{array}{c}
Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \\
\downarrow \quad \downarrow \quad \downarrow \\
\text{3 blocs} \\
= \\
\text{3 sous espaces vectoriels orthogonaux}
\end{array}
$$

On vérifie bien l'orthogonalité du plan.

Remarque: $\tilde{X}'\tilde{X}$ est facile à inverser.

Remarque pour le formateur:
c'est parce que la matrice $\tilde{X}'\tilde{X}$ est bloc-diagonale (parce que les trois sous-espaces vectoriels sont orthogonaux) que la décomposition des sommes des carrés est unique (cf. cours jour 3:ANOVA).

Rendre X de plein rang
$\Longrightarrow X'X$ inversible

Solution 1	Supprimer des colonnes de X $\Longleftrightarrow$ contraintes $\begin{cases} \alpha_1 & = 0 \\ \beta_1 & = 0 \end{cases}$ $\Longrightarrow$ paramétrisation fixée
Solution 2	Recentrer la matrice X $\Longleftrightarrow$ contraintes $\begin{cases} \alpha_1 + \alpha_2 & = 0 \\ \beta_1 + \beta_2 + \beta_3 & = 0 \end{cases}$ $\Longrightarrow$ paramétrisation fixée

Tableau récapitulatif:

deux solutions, qui correspondent à

- Deux paramétrisations,
- Deux systèmes de contraites.

Exo 1 : 1 donnée manquante

	B	1	2	3
A	1	-	- -	- -
	2	- -	- -	- -

11 données

$$\widetilde{X}' \quad . \quad \widetilde{X}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 1 & 1 & -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & -1 & -1 & 1 & 1 & 0 & 0 & -1 & -1 & 1 & 1 \end{pmatrix} . \begin{pmatrix} 1 & -1 & -1 & 0 \\ 1 & -1 & 0 & -1 \\ 1 & -1 & 0 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

11 lignes

$\widetilde{X}'\widetilde{X} =$

11	0	0	0
0	11	-1	0
0	-1	7	4
0	0	4	8

bloc diagonale : FAUX , plan non orthogonal

Supposons qu'il manque une observation, par exemple la première.

Remarque:
Les vecteurs colonne de $\tilde{X}$ ne sont plus orthogonaux d'un bloc à l'autre.

Ce qui se traduit sur la matrice $\tilde{X}'\tilde{X}$ par le fait qu'elle n'est plus bloc-diagonale.

Remarque pour le formateur:
On n'a donc pas unicité de la décomposition des sommes des carrés (cf. jour 3: ANOVA). (La décomposition dépend du choix des contraintes).

Exo 2

	B	1	2	3
A	1	-	-	-
	2	- -	- -	- -

Plan orthogonal ?

$$X = \begin{array}{c} \begin{array}{cccccc} 1 & A_1 & A_2 & B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ & & & & & \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

$\tilde{X}$ obtenue par les transformations : $\begin{cases} \overrightarrow{\mathbf{1}} & \longrightarrow \overrightarrow{\mathbf{1}} \\ \overrightarrow{A_2} & \longrightarrow \overrightarrow{A_2} - 2\,\overrightarrow{A_1} \\ \overrightarrow{B_2} & \longrightarrow \overrightarrow{B_1} - \overrightarrow{B_3} \\ \overrightarrow{B_3} & \longrightarrow \overrightarrow{B_2} - \overrightarrow{B_3} \end{cases}$

$\overrightarrow{\mathbf{1}}$ *et* $\overrightarrow{A_2} - 2\,\overrightarrow{A_1}$ *sont orthogonaux alors que* $\overrightarrow{\mathbf{1}}$ *et* $\overrightarrow{A_2} - \overrightarrow{A_1}$ *ne le sont pas*

TRANSPARENT FACULTATIF

$$\widetilde{X}' \cdot \widetilde{X}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -2 & -2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 1 & -1 & 0 & 0 & 1 & 1 & -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -2 & 1 & 0 \\ 1 & -2 & 0 & 1 \\ 1 & -2 & -1 & -1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

$\widetilde{X}' \widetilde{X} =$

9	0	0	0
0	18	0	0
0	0	6	3
0	0	3	6

bloc diagonale

1 bloc	$\longrightarrow$ **1**
1 bloc	$\longrightarrow$ A
1 bloc	$\longrightarrow$ B

1 bloc pour chaque terme du modèle

$$\begin{array}{ccccccccc} Y_{ijk} & = & \mu & + & \alpha_i & + & \beta_j & + & \varepsilon_{ijk} \\ & & \downarrow & & \downarrow & & \downarrow & & \end{array}$$

$$\text{3 blocs} = \text{3 s.e.v. orthogonaux}$$

plan orthogonal

TRANSPARENT FACULTATIF

On peut se rendre compte, sur la matrice $\tilde{X}$ que tout vecteur colonne d'un bloc est orthogonal à tout vecteur colonne d'un autre bloc.

Ce qui se traduit par le fait que la matrice $\tilde{X}'\tilde{X}$ est bloc-diagonale.

Le plan est orthogonal.

Plan

		B 1	2	3	
A	1	1	1	1	3
	2	2	2	2	6
		3	3	3	9

orthogonal
car :

nb répétitions proportionnel
- d'une ligne à l'autre
- d'une colonne à l'autre

$$n_{ij} = \frac{n_{i+}.n_{+j}}{n_{++}}$$

$$n_{i+} = \sum_j n_{ij}$$
$$n_{+j} = \sum_i n_{ij}$$
$$n_{++} = \sum_{i,j} n_{ij}$$

Dans ce cas, on utilise les contraintes :

$$
\begin{cases}
\sum\limits_{i} n_{i+}\, \alpha_i & = 0 \\
\sum\limits_{j} n_{+j}\, \beta_j & = 0
\end{cases}
$$

Voici le plan **que l'on vient d'étudier**.

Remarque: Dans certains livres, on trouve ce choix de contraintes:

$$\begin{cases} \sum_i n_{i+}\,\alpha_i = 0 \\ \sum_j n_{+j}\,\beta_j = 0 \end{cases}$$

Il a l'avantage de rendre le plan orthogonal, mais il a l'inconvénient de donner des hypothèses testées peu interprétables.
Jean-Marc Azaïs préconise le choix de contraintes:

$$\begin{cases} \sum_i \alpha_i = 0 \\ \sum_j \beta_j = 0 \\ \left.\begin{array}{l} \sum_i \gamma_{ij} = 0 \\ \sum_j \gamma_{ij} = 0 \end{array}\right) \text{si interaction.} \end{cases}$$

(C'est-à-dire le choix de la somme des carrés de type III).
Dans ce cas, le plan n'est pas orthogonal, mais les hypothèses testées sont interprétables.
Exemple de vérification de $n_{ij} = \frac{n_{i+}.n_{+j}}{n_{++}}$ pour $i = 2, j = 1$:

$$n_{21} = \frac{6 \times 3}{9} = 2$$

$$\begin{cases} 3\alpha_1 + 6\alpha_2 & = 0 \\ 3\beta_1 + 3\beta_2 + 3\beta_3 & = 0 \end{cases} \text{ soit } \begin{cases} \alpha_1 + 2\alpha_2 & = 0 \\ \beta_1 + \beta_2 + \beta_3 & = 0 \end{cases}$$

$$\overrightarrow{X\theta} = \mu.\ \overrightarrow{\mathbf{1}}\ +\alpha_1.\ \overrightarrow{A_1}\ +\alpha_2.\ \overrightarrow{A_2}\ +\beta_1.\ \overrightarrow{B_1}\ +\beta_2.\ \overrightarrow{B_2}\ +\beta_3.\ \overrightarrow{B_3}$$

or contraintes $\begin{cases} \alpha_1 & = -2\alpha_2 \\ \beta_3 & = -(\beta_1 + \beta_2) \end{cases}$

D'où :

$$\begin{aligned} \overrightarrow{X\theta} &= \mu.\ \overrightarrow{\mathbf{1}}\ -2\alpha_2.\ \overrightarrow{A_1}\ +\alpha_2.\ \overrightarrow{A_2}\ +\beta_1.\ \overrightarrow{B_1}\ +\beta_2.\ \overrightarrow{B_2} - (\beta_1 + \beta_2)\ .\ \overrightarrow{B_3} \\ &= \mu.\ \overrightarrow{\mathbf{1}}\ +\alpha_2.\left(\overrightarrow{A_2}\ -2\ \overrightarrow{A_1}\right) + \beta_1.\left(\overrightarrow{B_1} - \overrightarrow{B_3}\right) + \beta_2.\left(\overrightarrow{B_2} - \overrightarrow{B_3}\right) \\ &= \overrightarrow{\tilde{X}\theta} \end{aligned}$$

Donc, en transformant X en $\tilde{X}$, on a bien utilisé les contraintes :

$$\begin{cases} \sum_i n_{i+}\,\alpha_i & = 0 \\ \sum_j n_{+j}\,\beta_j & = 0 \end{cases}$$

TRANSPARENT FACULTATIF

Montrons que:
Transformer X en $\tilde{X}$ comme on l'a fait revient bien à utiliser les contraintes données au transparent précédent.