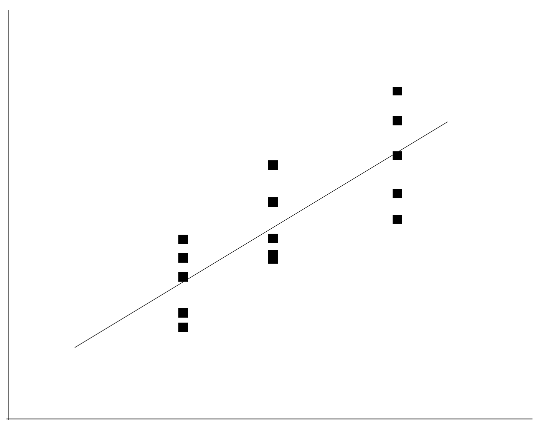
GÉNÉRALISATION À L'ENSEMBLE DU MODÈLE LINÉAIRE

- **en ANOVA**

$$M \ : \ Y_{ij} = \mu_i + \varepsilon_{ij}$$

exemple : réponse à une dose d'engrais X_i



Est-ce une droite ?

$$M_0 \ : \ Y_{ij} = \alpha + \beta X_i + \varepsilon_{ij}$$

$$\Longleftrightarrow \begin{cases} \mu_1 = \alpha + \beta X_1 \\ \mu_2 = \alpha + \beta X_2 \\ \mu_3 = \alpha + \beta X_3 \end{cases}$$

Quand le facteur représente un critère de classification de nature quantitative.

Revient à savoir s'il y a une relation linéaire entre μ_1, μ_2, μ_3.

facultatif : pour les stagiaires qui connaissent la méthode des contrastes, on peut dire (?) que cela revient à des tests de sous-modèles correspondants à des combinaisons linéaires d'effets de facteurs.

- **en RÉGRESSION**

$$M \ : \ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

– **Egalité de 2 coefficients ?**

$$H_0 \ : \ \beta_3 = \beta_2$$

$$M_0 \ : \ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 Z_i + \varepsilon_i$$

avec $Z_i = X_{2i} + X_{3i}$

– **coefficient égal à une valeur donnée ?**

$$H_0' \ : \ \beta_1 = 1$$

$$M_0' \ : \ W_i = \beta_0 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

avec $W_i = Y_i - X_{1i}$

Dans M_0, on crée un nouveau régresseur $X_2 + X_3$.

Dans M_0', on crée une nouvelle variable expliquée $Y - X_1$.

– Comparaison de droites de régression

$$M \ : \ \begin{cases} Y_{1j} = \beta_{10} + \beta_{11} X_{1j} + \varepsilon_{1j} \\ Y_{2j} = \beta_{20} + \beta_{21} X_{2j} + \varepsilon_{2j} \end{cases}$$

en introduisant la variable Z_{ij} telle que
$\forall j \ Z_{1j} = 0$ **et** $Z_{2j} = 1$ **:**

$$M \Longleftrightarrow Y_{ij} = \beta_0 + \gamma Z_{ij} + \beta_1 X_{ij} + \delta Z_{ij} X_{ij} + \varepsilon_{ij}$$

$$\begin{cases} \beta_{10} = \beta_0 \\ \beta_{20} = \beta_0 + \gamma \end{cases} \begin{cases} \beta_{11} = \beta_1 \\ \beta_{21} = \beta_1 + \delta \end{cases}$$

$*$ **droites parallèles ?**

$$H_0 : \beta_{11} = \beta_{21} \Longleftrightarrow \delta = 0$$

$$M_0 \ : \ Y_{ij} = \beta_0 + \gamma Z_{ij} + \beta_1 X_{ij} + \varepsilon_{ij}$$

$*$ **même droite ?**

$$H_0' : \begin{cases} \beta_{10} = \beta_{20} \\ \beta_{11} = \beta_{21} \end{cases} \Longleftrightarrow \begin{cases} \gamma = 0 \\ \delta = 0 \end{cases}$$

$$M_0' \ : \ Y_{ij} = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}$$

Montrer comment on retrouve les coefficients initiaux :
si $i = 1$ alors $Z_{i,j} = 0$

$$y_{1j} = \beta_0 + \beta_1 X_{1j} + \varepsilon_{1j} \qquad (1)$$

donc $\beta_{10} = \beta_0$ et $\beta_{11} = \beta_1$

si $i = 2$ alors $Z_{i,j} = 1$

$$\begin{aligned} y_{2j} &= \beta_0 + \gamma \times 1 + \beta_1 X_{2j} + \delta_1 \times 1\, X_{2j} + \varepsilon_{2j} && (2) \\ &= (\beta_0 + \gamma) + (\beta_1 + \delta) X_{2j} + \varepsilon_{2j} && (3) \end{aligned}$$

d'où $\beta_{20} = \beta_0 + \gamma$ et $\beta_{21} = \beta_1 + \delta$

Le modèle M_0 revient à un modèle de *covariance* (vu lors de la journée "analyse de variance").

- **ANALYSE DE COVARIANCE**

$$M \ : \ Y_{ij} = \mu_i + \beta X_{ij} + \varepsilon_{ij}$$

  - **pas d'effet de la covariable ?**

$$H_0 : \beta = 0$$

$$M_0 \ : \ Y_{ij} = \mu_i + \varepsilon_{ij}$$

  - **pas d'effet du facteur ?**

$$H'_0 : \forall i \ \mu_i = \mu$$

$$M'_0 \ : \ Y_{ij} = \mu + \beta X_{ij} + \varepsilon_{ij}$$

Exemple :
On étudie la prise de poids de lapins après engraissement selon plusieurs régimes :

$$
\begin{aligned}
Y &= poids\ apres\ experience & (4)\\
i &= regime\ alimentaire & (5)\\
covariable\ X &= poids\ initial\ de\ l'animal & (6)
\end{aligned}
$$

M_0 <=> ANOVA

exemple : le poids initiale n'a pas d'effet

M_0' <=> régression

exemple : le régime n'a pas d'effet

REMARQUES ET MISES EN GARDE

- **Plusieurs modèles peuvent décrire valablement un même lot de données.**

- **Pas d'équivalence automatique en régression multiple entre significativité des tests**
  modèle (Fisher) $\not\equiv$ *coefficients du modèle (Student)*

- **Colinéarité des régresseurs**
  $\Longrightarrow$ *matrice* $X'X$ *non inversible et indétermination sur l'estimation des paramètres*

- **Régresseurs non orthogonaux**
  $\Longrightarrow$ *instabilité des estimations (coefficients partiels)*

Choix difficile du modèle : l'expérimentateur choisira la plupart du temps selon des critères non-statistiques (compromis précision-coût).

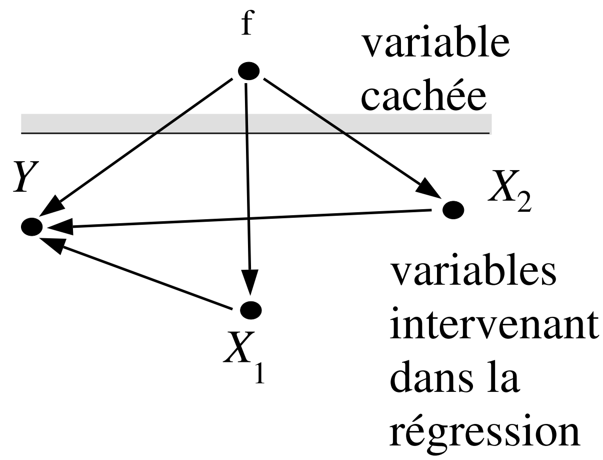
On peut avoir un modèle significatif et peu ou pas de coefficient significatif et inversement.

Quand il existe une combinaison linéaire liant des régresseurs entre eux, il faut analyser la structure de la matrice X et tenter de supprimer la colinéarité en ôtant des régresseurs.

Lorsque les régresseurs sont plus ou moins corrélés, les estimations des coefficients ne sont pas indépendants entre eux, il faut les interpréter dans leur ensemble.

facultatif : citer la méthode de la régression orthogonalisée qui consiste à remplacer les régresseurs par des variables globalement équivalentes orthogonales.

- **Variables sous-jacentes**



Attention au schéma causal !

- **Interprétation des résultats d'analyse en relation avec le plan d'expérience**

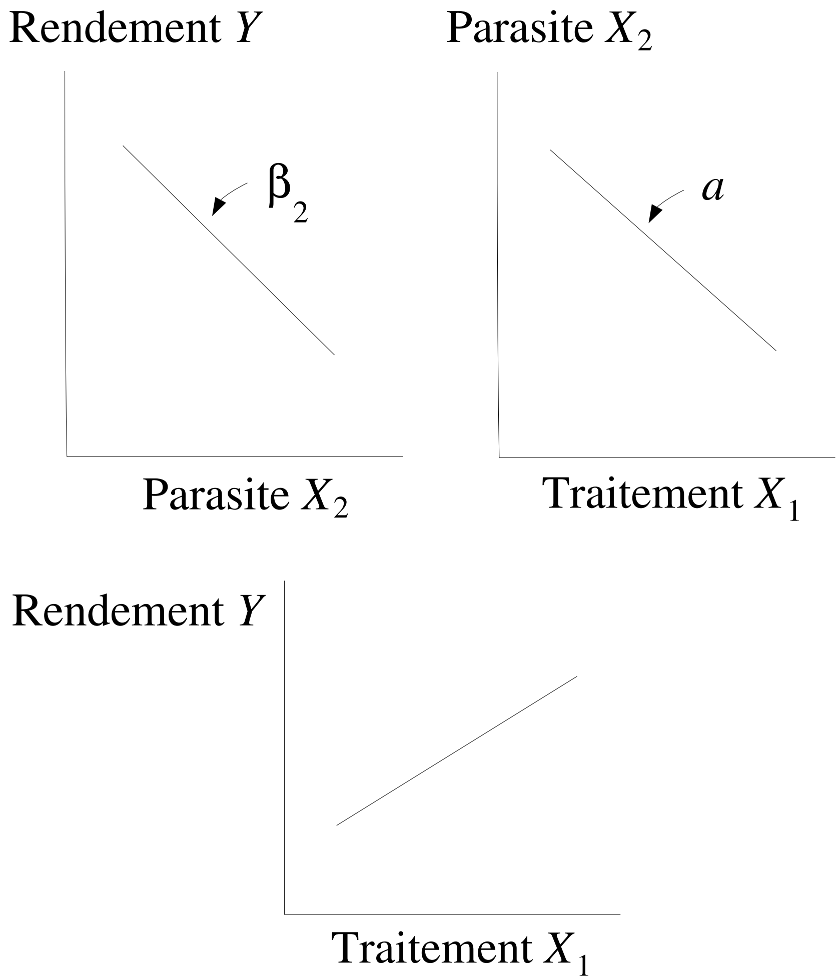
si possible :
planifier les expériences $\Rightarrow$ *orthogonalité ,*
indépendance

Il arrive très fréquemment que les variables mesurées dont on dispose pour le modèle de régression (taille, poids, etc...) soient diffèrents aspects de la même variable qui est immensurable. Dans l'exemple de la moutarde, la "croissance" est en fait la variable cachée derrière les variables mesurées.

facultatif : exemple de l'attaque parasitaire, voir transparent suivant.

On a vu que l'inférence dépend de la structure de la matrice X (matrice d'expérience).
Il est important de s'assurer que l'utilisation du modèle est correcte pour le plan d'expérience en question -postulats respectés,etc..- .
Il faut également être conscient des confusions d'effets entraînées par certains plans.

illustration d'un schéma causal :
attaque parasitaire d'une culture végétale



$$\hat{Y} = b_1 X_1$$

b_1 est une estimation de $a\beta_2$

Soit une culture pour laquelle il existe un parasite dont l'effet peut être diminué grâce à un traitement chimique (figure de droite).

Supposons que le rendement Y de la culture ne soit pas directement influencé par le traitement X_1, par contre le parasite X_2 diminue le rendement (figure de gauche).

Si nous calculons la régression de Y sur X_1, nous la trouvons significative (figure du bas), or $\boldsymbol{b}_1$ ne peut être interprété comme l'effet sur Y du changement d'une unité de X_1. Dans des conditions parasitaires différentes ces résultats ne sont plus valables.