COMMENT ?

- **Restreindre un modèle**

- **Tester des hypothèses linéaires sur les paramètres d'un modèle**

$\Longrightarrow$ $\boxed{\textbf{TESTS DE SOUS-MODÈLES}}$

Un sous-modèle est un cas particulier d'un modèle général contenant moins de paramètres que celui-ci.

Jusqu'à présent, on a vu comment tester à partir d'un modèle :

- la nullité d'un coefficient : test de Student,
- la nullité de tous les coefficients des régresseurs : test F du modèle,

Maintenant,

- Peut-on, à partir d'un modèle valide, supprimer des régresseurs, pour des raisons expérimentales -mesures difficiles à réaliser, etc...- (***la valeur du test de Student n'est pas un critère de suppression, même si celui-ci n'est pas significatif pour un régresseur, ce n'est pas une raison pour le supprimer si celui-ci est toujours mesuré***) ?
- comment tester la nullité d'un ensemble de coef-ficients ou toute autre combinaison linéaire (si ça a un sens, ex : contraintes à priori) ?

DÉMARCHE DU TEST DE SOUS-MODELE

- **Modèle général**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$$

une précédente analyse a montré que ce modèle est valide

- **Hypothèses à tester**

H_0 **:** $(\beta_3 = \beta_4 = 0)$
H_1 **:** **(**$\beta_3 \neq 0$ **et/ou** $\beta_4 \neq 0$**)**

- **Sous-modèle correspondant à** H_0

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon'_i$$

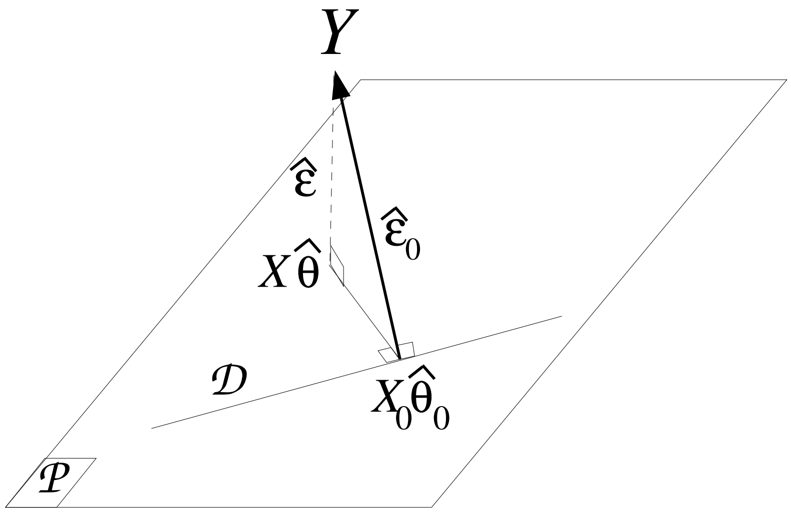
Dans le sous-modèle (appelé également modèle emboîté), la quantité ε' contient ε + la variabilité expliquée par X_3 et X_4.

• Les Analyses de Variances

modèle	Origine de la variation	Sommes des carrés	d.l.
Modèle complet	**Reg.** X_1, X_2, X_3, X_4	SCM_M	4
M	**variation résiduelle**	SCR_M	$n-4-1$
Sous-modèle	**Reg.** X_1, X_2	SCM_{M_0}	2
M_0	**variation résiduelle**	SCR_{M_0}	$n-2-1$
M **et** M_0	**variation totale**	SCT	$n-1$

Il faut bien sûr contrôler la structure des résidus.

• Représentation géométrique



$\mathcal{P}$: espace vectoriel $X\Theta$ de dimension p représentant M
$\mathcal{D}$: sous-espace vectoriel $X\Theta_0$ de dimension q représentant M_0 (avec $q < p$)

• Ecart entre les deux modèles

$$\Delta = SCR_{M_0} - SCR_M$$

avec $(n-q-1)-(n-p-1)=p-q$ *degré de liberté*

Montrer sur le graphe les différentes quantités :

$$
\begin{aligned}
SCR_M &= \sum \|Y \overset{\rightarrow}{-} X\hat{\theta}\|^2 \\
SCR_{M_0} &= \sum \|Y \overset{\rightarrow}{-} X_0\hat{\theta}_0\|^2 \\
&= \sum(\|Y \overset{\rightarrow}{-} X\hat{\theta}\|^2 + \|X\hat{\theta} \overset{\rightarrow}{-} X_0\hat{\theta}_0\|^2) \\
&= SCR_M + \sum \|X\theta \overset{\rightarrow}{-} X_0\hat{\theta}_0\|^2
\end{aligned}
$$

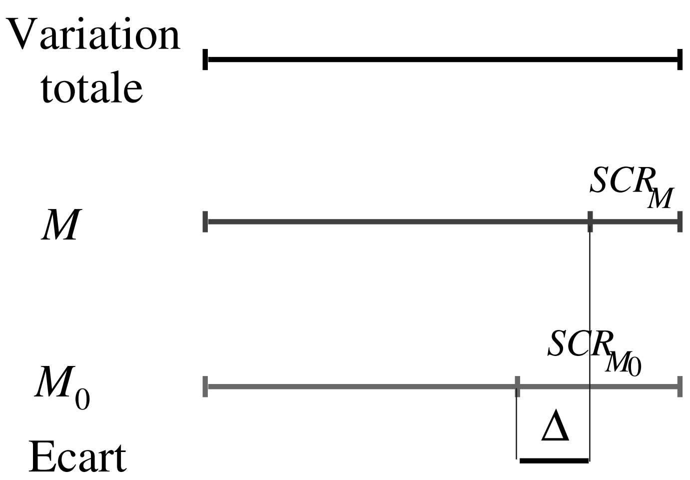
d'où

$$\Delta = \sum \|X\theta \overset{\rightarrow}{-} X_0\widehat{\theta_0}\|^2 = SCR_{M_0} - SCR_M$$

faire le calcul des degrés de liberté :

$$
\begin{aligned}
Y &: n-1 \\
X\hat{\theta} &: p \\
SCR_M &: n-p-1 \\
X_0\hat{\theta}_0 &: q \\
SCR_{M_0} &: n-q-1
\end{aligned}
$$

- **Autre représentation graphique**



- **Test du sous-modèle**

$$\frac{(SCR_{M_0} - SCR_M)/(p-q)}{SCR_M/(n-p-1)} \overset{H_0}{\sim} F(p-q, n-p-1)$$

Le test consiste à comparer Δ à la variance estimée par $SCR_M/(n-p-1)$

Si les résidus sont distribués normalement.

On rejette H_0 au risque α lorsque cette statistique est supérieure au $1-\alpha$ quantile de F à (p-q),(n-p-1) ddl.

EXEMPLE

- **Modèle général**

$M : RC = \beta_0 + \beta_1 LT + \beta_2 HF + \beta_3 PR + \beta_4 PA + \beta_5 FE + \varepsilon$

Source	DF	Sum of squares	Mean Square	F Value	Prob>F
Model	5	65.21663	13.04333	50.539	0.0001
Error	25	6.45208	0.25808		
Total	30	71.66871			

- **Hypothèses à tester**

$H_0 : \beta_2 = \beta_4 = \beta_5 = 0$

- **Sous-modèle correspondant à H_0**

$M_0 : RC = \beta_0 + \beta_1 LT + \beta_3 PR + \varepsilon$

Source	DF	Sum of squares	Mean Square	F Value	Prob>F
Model	2	65.53090	31.26545	95.803	0.0001
Error	28	9.13781	0.32635		
Total	30	71.66871			

$$F = \frac{(9.14 - 6.45)/(28 - 25)}{6.45/25} = 3.47$$

à comparer à $F_{3,25,5\%} = 3.38 \Rightarrow$ *on rejette H_0 au risque de 5% de se tromper*

On reprend l'exemple de la moutarde. L'expérimentateur aimerait, pour une expérience ultérieure se dispenser de la mesure du potentiel hydrique qui est lourde à réaliser, du comptage des feuilles, ainsi que de la pesée de la partie aérienne qui nécessite de nombreuses manipulations.

Conclusion : le sous-modèle ne convient pas.

- **Modèle général**

$M : RC = \beta_0 + \beta_1 LT + \beta_2 HF + \beta_3 PR + \beta_4 PA + \beta_5 FE + \varepsilon$

Source	DF	Sum of squares	Mean Square	F Value	Prob>F
Model	5	65.21663	13.04333	50.539	0.0001
Error	25	6.45208	0.25808		
Total	30	71.66871			

- **Hypothèses à tester**

$H_0' : \beta_2 = \beta_5 = 0$

- **Sous-modèle correspondant à H_0**

$$M_0' : RC = \beta_0 + \beta_1 LT + \beta_3 PR + \beta_4 PA + \varepsilon$$

Source	DF	Sum of squares	Mean Square	F Value	Prob>F
Model	3	64.41092	21.47031	79.873	0.0001
Error	27	7.25779	0.26881		
Total	30	71.66871			

$$F = \frac{(7.26 - 6.45)/(27 - 25)}{6.45/25} = 1.57$$

à comparer à $F_{3,25,5\%} = 2.99 \Rightarrow$ *on ne peut rejeter* H_0

L'expérimentateur aimerait quand même pouvoir supprimer la mesure du potentiel hydrique et le comptage des feuilles => autre sous-modèle.

On accepte donc le sous-modèle.

- **Estimations des paramètres**

$$M : RC = \beta_0 + \beta_1 LT + \beta_2 HF + \beta_3 PR + \beta_4 PA + \beta_5 FE + \varepsilon$$

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> |T|
INTERCEP	1	-0.428901	0.40886893	-1.049	0.3042
LT	1	0.013582	0.00794030	1.711	0.0995
HF	1	0.003234	0.00210214	1.539	0.1365
PR	1	0.002889	0.00058155	4.967	0.0001
PA	1	-0.000285	0.00013538	-2.107	0.0453
FE	1	-0.054721	0.07386775	-0.741	0.4657

$$M_0' : RC = \beta_0 + \beta_1 LT + \beta_3 PR + \beta_5 FE + \varepsilon$$

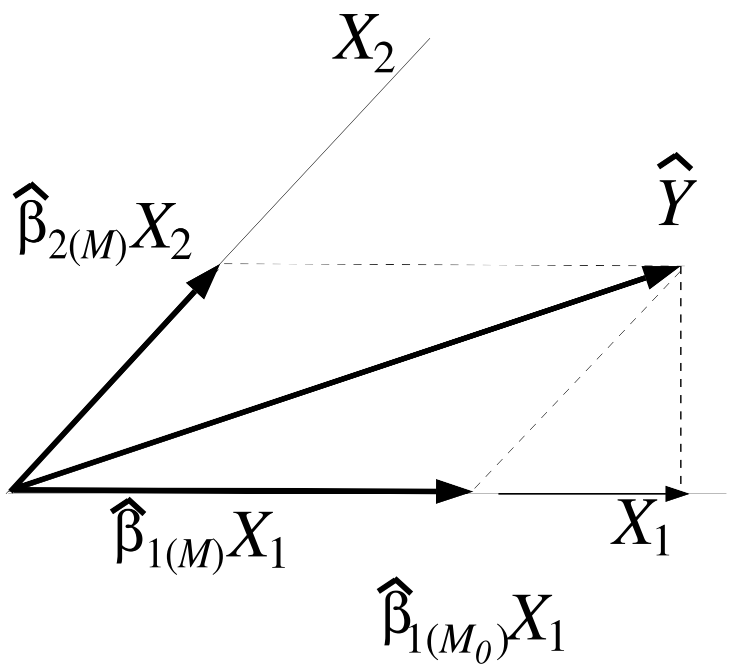
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> |T|
INTERCEP	1	-0.344535	0.35081919	-0.982	0.3348
LT	1	0.009461	0.00774552	1.221	0.2325
PA	1	-0.000347	0.00013133	-2.645	0.0135
PR	1	0.003381	0.00034183	9.891	0.0001

les coefficients estimés ne sont pas identiques *car les régresseurs ne sont pas orthogonaux*

illustration géométrique de la non-orthogonalité

$$M : Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$M_0 : Y = \beta_1 X_1 + \varepsilon$$



$$\hat{\beta}_{1(M_0)} \neq \hat{\beta}_{1(M)}$$

Soit 2 régresseurs non-orthogonaux X_1 et X_2.

Les estimations des coefficients ne sont pas indépendantes entre elles. En effet si l'on supprime un régresseur du modèle (par exemple X_2), l'estimation du paramètre restant est différente.

Dans de telles situations, les estimations ne peuvent être interprétées que les unes par rapport aux autres, et non indépendemment les unes des autres.

- **Modèle général**

  $M : RC = \beta_0 + \beta_1 LT + \beta_2 HF + \beta_3 PR + \beta_4 PA + \beta_5 FE + \varepsilon$

Source	DF	Sum of squares	Mean Square	F Value	Prob>F
Model	5	65.21663	13.04333	50.539	0.0001
Error	25	6.45208	0.25808		
Total	30	71.66871			

- **Hypothèses à tester**

  $H_0'' : \beta_4 = \beta_3$

- **Sous-modèle correspondant à H_0''**

  $M_0'' : RC = \beta_0 + \beta_1 LT + \beta_2 HF + \beta_3 (PR + PA) + \beta_5 FE + \varepsilon$

Source	DF	Sum of squares	Mean Square	F Value	Prob>F
Model	4	58.89049	17.72262	29.956	0.0001
Error	26	12.77822	0.49147		
Total	30	71.66871			

$$F = \frac{(12.78 - 6.45)}{6.45/25} = 24.53$$

à comparer à $F_{1,25,5\%} = 4.24 \Rightarrow$ *on rejette* H_0'' *au risque de 5% de se tromper*

EXERCICE :

L'expérimentateur voudrait peser la plante entière PT au lieu de séparer la partie aérienne des racines.

Quelles sont les hypothèses à tester et le sous-modèle correspondant ?

on fournit les tableaux d'analyse de variance des 2 modèles

Calculer la statistique et réaliser le test.