Tableau "Modèle"

General Linear Models Procedure

Class Level Information

Class	Levels	Values
TESTEUR	2	1 2
ORIGINE	3	1 2 3

Number of observations in data set = 12

General Linear Models Procedure

Dependent Variable : HUILE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	47.18144167	9.43628833	6.18	0.0233
Error	6	9.16665000	1.52777500		
Corrected Total	11	56.34809167			

R-Square	C.V.	Root MSE	HUILE Mean
0.837321	2.671010	1.236032	46.27583

$$R^2 = \frac{SCE_{M_0} - SCE_{M_3}}{SCE_{M_0}}$$

Source			
Model	IJ-1	$SCE_{M_0} - SCE_{M_3}$	$\dfrac{\left(SCE_{M_0} - SCE_{M_3}\right) / IJ-1}{SCE_{M_3} / N-IJ}$
Error	N-IJ	SCE_{M_3}	
Corrected Total	N-1	SCE_{M_0}	

Le test de F du modèle étant significatif, cela nous autorise à rechercher quelles sont les sources de variation qui structurent les données.

FPSTAT INRA IV-Analyse de Variance-25

Tableau "Facteurs"

Dependent Variable : HUILE

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TESTEUR	1	9.48740833	9.48740833	6.21	0.0470
ORIGINE	2	33.74581667	16.87290833	11.04	0.0097
TESTEUR*ORIGINE	2	3.94821667	1.97410833	1.29	0.3415

Dependent Variable : HUILE

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TESTEUR	1	9.48740833	9.48740833	6.21	0.0470
ORIGINE	2	33.74581667	16.87290833	11.04	0.0097
TESTEUR*ORIGINE	2	3.94821667	1.97410833	1.29	0.3415

TESTEUR	I-1	$SCE_{M_0} - SCE_{M_1}$	$\dfrac{\left(SCE_{M_0} - SCE_{M_1}\right) / I-1}{SCE_{M_3} / N-IJ}$
ORIGINE	J-1	$SCE_{M_1} - SCE_{M_2}$	$\dfrac{\left(SCE_{M_1} - SCE_{M_2}\right) / J-1}{SCE_{M_3} / N-IJ}$
TESTEUR*ORIGINE	(I-1)(J-1)	$SCE_{M_2} - SCE_{M_3}$	$\dfrac{\left(SCE_{M_2} - SCE_{M_3}\right) /(I-1)(J-1)}{SCE_{M_3} / N-IJ}$

Faire remarquer que le premier facteur testé est celui qui est déclaré, dans le modèle, le premier. La sortie correspond à l'ordre déclaré dans le modèle.

Les degrés de liberté pour l'interaction se calculent de la façon suivante :

$$\mathrm{IJ} - (\mathrm{I} + \mathrm{J} - 1) = (\mathrm{I} - 1)\,(\mathrm{J} - 1)$$

IJ : nombre de paramètres irréductibles pour le modèle interactif $(\mathrm{M}_3 \text{ écriture } \mu_{\mathrm{ij}})$

I+J-1 : nombre de paramètres irréductibles du modèle additif (M_2) ; *l'écriture irréductible n'est pas possible* mais dans la matrice X, $\sum \mathrm{A_i} = \sum \mathrm{B_j} = \mathbb{1}$ il existe une relation liant les vecteurs $\mathrm{A_i}$ et $\mathrm{B_j}$ de la matrice, d'où I (pour le nombre de vecteurs $\mathrm{A_i}$) + J (pour le nombre de vecteurs $\mathrm{B_j}$) − 1 (pour la relation entre les $\mathrm{A_i}$ et $\mathrm{B_j}$). *Les notations correspondent au chapitre III*

On peut aussi dire que l'interaction se définissant comme un écart au modèle additif, on étudie l'effet de la combinaison des deux facteurs $(\mathrm{M}_3/\mathrm{M}_0)$ après avoir enlevé l'effet du premier facteur $(\mathrm{M}_1/\mathrm{M}_0)$ et celui du deuxième facteur $\left(\mathrm{M}'_1/\mathrm{M}_0\right)$.

Soit pour les degrés de liberté :

$$(\mathrm{IJ} - 1) - (\mathrm{I} - 1) - (\mathrm{J} - 1) = (\mathrm{I} - 1)(\mathrm{J} - 1)$$

Fichier des données et Commandes SAS

```
1 1 43.54
1 1 45.30
1 2 44.25
1 2 42.55
1 3 47.28
1 3 49.40
2 1 47.21
2 1 47.73
2 2 44.34
2 2 46.49
2 3 47.75
2 3 49.47
```

```
data trv ;
infile 'tourn' ;
input testeur origine huile ;
run ;
proc glm ;
class testeur origine ;
model huile = origine testeur testeur*origine ;
run ;
```

M_3 (testeur + origine + interaction testeur x origine)

$\mu_n = \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$

M_2 (testeur + origine)

$\mu_n = \mu + \alpha_i + \beta_j$

M'_1 (origine)

$\mu_n = \mu_j = \mu + \beta_j$

M_0

$\mu_n = \mu$

FPSTAT INRA IV-Analyse de Variance-27

Sortie SAS

General Linear Models Procedure

Class Level Information

Class	Levels	Values
Testeur	2	1 2
Origine	3	1 2 3

Number of observations in data set = 12

General Linear Models Procedure

Dependent Variable : HUILE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	47.18144167	9.43628833	6.18	0.0233
Error	6	9.16665000	1.52777500		
Corrected Total	11	56.34809167			

R-Square	C.V.	Root MSE	HUILE Mean
0.837321	2.671010	1.236032	46.27583

Dependent Variable : HUILE

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ORIGINE	2	33.74581667	16.87290833	11.04	0.0097
TESTEUR	1	9.48740833	9.48740833	6.21	0.0470
TESTEUR*ORIGINE	2	3.94821667	1.97410833	1.29	0.3415

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ORIGINE	2	33.74581667	16.87290833	11.04	0.0097
TESTEUR	1	9.48740833	9.48740833	6.21	0.0470
TESTEUR*ORIGINE	2	3.94821667	1.97410833	1.29	0.3415

Peut-être inutile, puisque l'on décompose le tableau

Tableau “Facteurs”

Dependent Variable : HUILE

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ORIGINE	2	33.74581667	16.87290833	11.04	0.0097
TESTEUR	1	9.48740833	9.48740833	6.21	0.0470
TESTEUR*ORIGINE	2	3.94821667	1.97410833	1.29	0.3415

Dependent Variable : HUILE

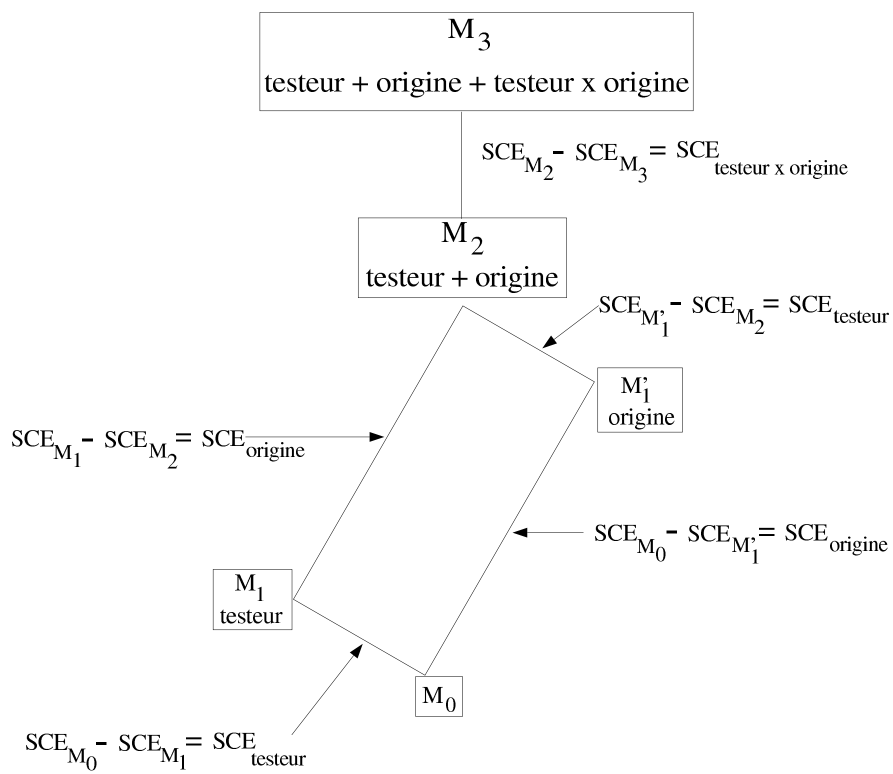
Source	DF	Type III SS	Mean Square	F Value	Pr > F
ORIGINE	2	33.74581667	16.87290833	11.04	0.0097
TESTEUR	1	9.48740833	9.48740833	6.21	0.0470
TESTEUR*ORIGINE	2	3.94821667	1.97410833	1.29	0.3415

ORIGINE	J-1	$SCE_{M_0} - SCE_{M'_1}$	$\dfrac{\left(SCE_{M_0} - SCE_{M'_1}\right) / J-1}{SCE_{M_3} / N-IJ}$
TESTEUR	I-1	$SCE_{M'_1} - SCE_{M_2}$	$\dfrac{\left(SCE_{M'_1} - SCE_{M_2}\right) / I-1}{SCE_{M_3} / N-IJ}$
TESTEUR*ORIGINE	(I-1)(J-1)	$SCE_{M_2} - SCE_{M_3}$	$\dfrac{\left(SCE_{M_2} - SCE_{M_3}\right) /(I-1)(J-1)}{SCE_{M_3} / N-IJ}$

On peut constater pour ce dispositif orthogonal que les sommes des carrés des écarts de type I sont égales à celles de type III. Les valeurs des sommes des carrés des écarts dans le modèle, sont indépendantes de l'ordre d'introduction des facteurs, c'est à dire :

$$\text{testeur} : SCE_{M_0} - SCE_{M_1} = SCE_{M'_1} - SCE_{M_2}$$
$$\text{origine} : SCE_{M_1} - SCE_{M_2} = SCE_{M_0} - SCE_{M'_1}$$

Remarque : L'interaction est toujours introduite dans le modèle après les deux facteurs.



Les deux analyses de variance faites successivement ainsi que les constatations sur les sommes des carrés des écarts peuvent être résumées par la figure présentée :

Pour le test des deux facteurs croisés, l'ensemble des deux analyses peut se résumer par un rectangle, dont la partie gauche correspond à la première analyse, la partie droite à la seconde. La valeur des sommes de carrés étant indépendante de l'ordre d'introduction des facteurs, chacun des côtés correspondant au test d'un même facteur a une longueur identique.

Remarque :

la figure tente de respecter le fait que la $\mathrm{SCE}_{\mathtt{testeur}}$ est plus petite que celle de l'origine. Elle est donc différente de la figure transparent IV–22 dont l'objectif est de présenter les cinq modèles envisagés, ceux qui sont emboités et ceux qui ne le sont pas.

Plan orthogonal

$$\text{SCE type I} = \text{SCE type III}$$

Une unique décomposition de la somme des carrés des écarts totale :

$$\text{SCE}_{\text{totale}} = \text{SCE}_{\text{testeur}} + \text{SCE}_{\text{origine}} + \text{SCE}_{\text{testeur x origine}} + \text{SCE}_{\text{résiduelle}}$$

FPSTAT INRA IV-Analyse de Variance-31

Plan orthogonal : que teste-t-on avec les tests F ?

Plan orthogonal $\Longleftrightarrow$ SCE type I = SCE type III

	F	H_0
origine	$\left\{\begin{array}{c}\dfrac{\left(SCE_{M_0}-SCE_{M_1'}\right)/\ J-1}{SCE_{M_3}/\ N-IJ}\\ = \\ \dfrac{\left(SCE_{M_1}-SCE_{M_2}\right)/\ J-1}{SCE_{M_3}/\ N-IJ}\end{array}\right.$	$\left\{\begin{array}{l}\forall\ j\\ \beta_j+\dfrac{\sum_i \gamma_{ij}}{I}=0\end{array}\right.$
testeur	$\left\{\begin{array}{c}\dfrac{\left(SCE_{M_0}-SCE_{M_1}\right)/\ I-1}{SCE_{M_3}/\ N-IJ}\\ = \\ \dfrac{\left(SCE_{M_1'}-SCE_{M_2}\right)/\ I-1}{SCE_{M_3}/\ N-IJ}\end{array}\right.$	$\left\{\begin{array}{l}\forall\ i\\ \alpha_i+\dfrac{\sum_j \gamma_{ij}}{J}=0\end{array}\right.$
testeur x origine	$\dfrac{\left(SCE_{M_2}-SCE_{M_3}\right)/\ IJ-1}{SCE_{M_3}/\ N-IJ}$	$\left\{\begin{array}{l}\forall\ (i,j)\\ \gamma_{ij}=0\end{array}\right.$

Les tests de F ne sont pas intrinsèques : l'effet d'un facteur est testé comme l'apport d'un modèle par rapport à un autre, aucun d'eux ne faisant intervenir l'interaction, contre la résiduelle d'un modèle faisant intervenir l'interaction. De ce fait <u>on ne teste pas un effet additif</u> α_i ou β_j mais <u>un effet principal</u> (l'effet additif et une fraction des interactions)

$$\text{ex}\left\{\begin{array}{ll}\beta_1+\dfrac{\sum_i \gamma_{i1}}{I} & \alpha_1+\dfrac{\sum_j \gamma_{1j}}{J}\\ \beta_2+\dfrac{\sum_i \gamma_{i2}}{I} & \alpha_2+\dfrac{\sum_j \gamma_{2j}}{J}\\ \vdots & \vdots\end{array}\right.$$

Analyse de Variance à deux facteurs cas non orthogonal

(pas de case vide)

3 variétés de carottes cultivées dans 2 sols

Y : nombre de jours avant germination

Données expérimentales

Y_{ijr}	Variété 1	Variété 2	Variété 3
Sol 1	6 10 11	13 15	14 22
Sol 2	12 19 15 18	31	18 9 12

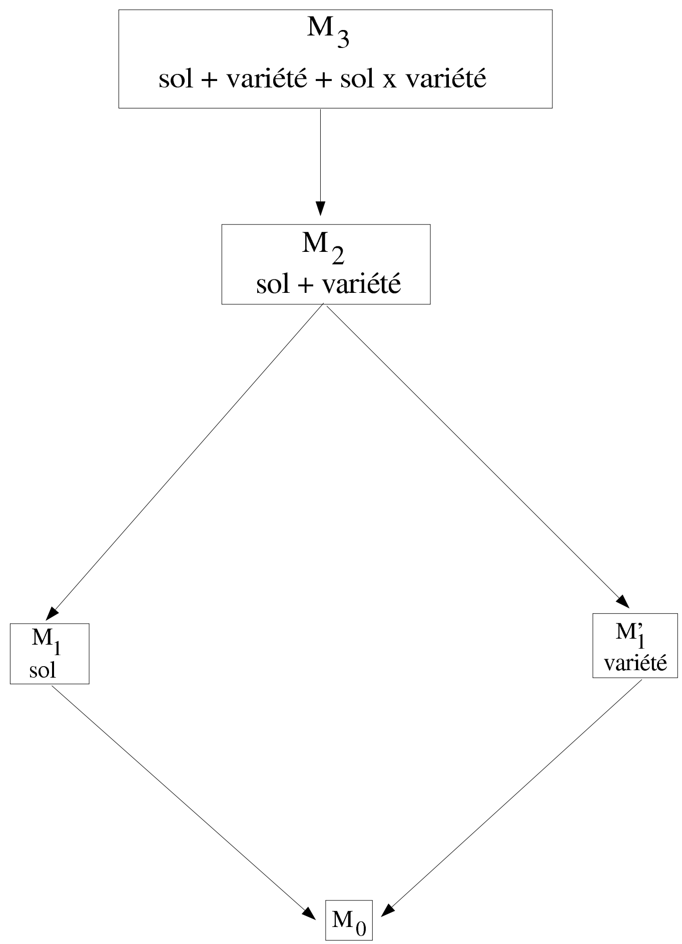
ex : dispositif déséquilibré

Facteur A = sol   I = 2 niveaux
Facteur B = variété   J = 3 niveaux

Fichier des données

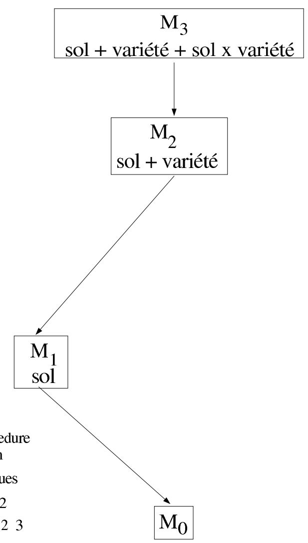
```
1 1 16
1 1 10
1 1 11
1 2 13
1 2 15
1 3 14
1 3 22
2 1 12
2 1 19
2 1 15
2 1 18
2 2 31
2 3 18
2 3  9
2 3 12
```


On verra plus tard que faire dans le cas de cases vides (cf. commentaire de IV-46)



Puisque 2 facteurs sont croisés, l'étude de la structuration des données fait intervenir 5 modèles.

```
data trv ;
infile 'solvar' ;
input sol variete germina ;
run ;
proc glm ;
class sol variete ;
model germina = sol variete  sol * variete ;
run ;
```



General Linear Models Procedure
Class Level Information

Class	Levels	Values
SOL	2	1 2
VARIETE	3	1 2 3

Number of observations in data set = 15

General Linear Models Procedure

Dependent Variable : GERMINA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	400.0000000	80.0000000	6.00	0.0103
Error	9	120.0000000	13.3333333		
Corrected Total	14	520.0000000			

R-Square	C.V.	Root MSE	GERMINA Mean
0.769231	24.34322	3.651484	15.00000

General Linear Models Procedure

Dependent Variable : GERMINA

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SOL	1	52.5000000	52.5000000	3.94	0.0785
VARIETE	2	124.7340426	62.3670213	4.68	0.0405
SOL*VARIETE	2	222.7659574	111.3829787	8.35	0.0089

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SOL	1	123.7714286	123.7714286	9.28	0.0139
VARIETE	2	192.1276596	96.0638298	7.20	0.0135
SOL*VARIETE	2	222.7659574	111.3829787	8.35	0.0089

FPSTAT INRA IV-Analyse de Variance-35

Tableau "Modèle"

General Linear Models Procedure

Class Level Information

Class	Levels	Values
SOL	2	1 2
VARIETE	3	1 2 3

Number of observations in data set = 15

General Linear Models Procedure

Dependent Variable : GERMINA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	400.0000000	80.0000000	6.00	0.0103
Error	9	120.0000000	13.3333333		
Corrected Total	14	520.0000000			

R-Square	C.V.	Root MSE	GERMINA Mean
0.769231	24.34322	3.651484	15.00000

$$r^2 = \frac{SCE_{M_0} - SCE_{M_3}}{SCE_{M_0}}$$

Source			
Model	IJ-1	$SCE_{M_0} - SCE_{M_3}$	$\dfrac{\left(SCE_{M_0} - SCE_{M_3}\right) / IJ-1}{SCE_{M_3} / N-IJ}$
Error	N-IJ	SCE_{M_3}	
Corrected Total	N-1	SCE_{M_0}	

FPSTAT INRA 1997 IV-Analyse de Variance-36

Tableau “Facteurs”

Dependent Variable : GERMINA

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SOL	1	52.5000000	52.5000000	3.94	0.0785
VARIETE	2	124.7340426	62.3670213	4.68	0.0405
SOL*VARIETE	2	222.7659574	111.3829787	8.35	0.0089

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SOL	1	123.7714286	123.7714286	9.28	0.0139
VARIETE	2	192.1276596	96.0638298	7.20	0.0135
SOL*VARIETE	2	222.7659574	111.3829787	8.35	0.0089

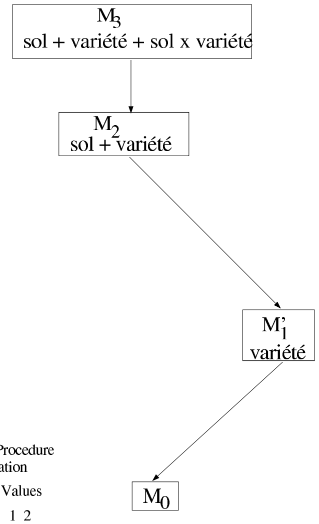
		Type I	
SOL	I-1	$SCE_{M_0} - SCE_{M_1}$	$\dfrac{\left(SCE_{M_0} - SCE_{M_1}\right) / I-1}{SCE_{M_3} / N-IJ}$
VARIETE	J-1	$SCE_{M_1} - SCE_{M_2}$	$\dfrac{\left(SCE_{M_1} - SCE_{M_2}\right) / J-1}{SCE_{M_3} / N-IJ}$
SOL*VARIETE	IJ-1	$SCE_{M_2} - SCE_{M_3}$	$\dfrac{\left(SCE_{M_2} - SCE_{M_3}\right) / IJ-1}{SCE_{M_3} / N-IJ}$

Pour les sommes de carrés des 2 facteurs, les sommes de type I diffèrent des sommes de type III.

Les sommes de carrés de type I pour l’interaction (effet de dernier ordre, introduit dans le modèle le plus complexe), sont égales à celles de type III.

Comme les sommes de type I et III diffèrent, les probabilités du F correspondant diffèrent. En conséquence, il peut y avoir discordance entre les probabilités des F pour les effets sol et variété.

```
data trv ;
infile 'solvar' ;
input sol variete germina ;
run ;
proc glm ;
class sol variete ;
model germina = variete sol  variete*sol ;
run ;
```



General Linear Models Procedure
Class Level Information

Class	Levels	Values
SOL	2	1 2
VARIETE	3	1 2 3

Number of observations in data set = 15

General Linear Models Procedure

Dependent Variable : GERMINA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	400.0000000	80.0000000	6.00	0.0103
Error	9	120.0000000	13.3333333		
Corrected Total	14	520.0000000			

R-Square	C.V.	Root MSE	GERMINA Mean
0.769231	24.34322	3.651484	15.00000

General Linear Models Procedure

Dependent Variable : GERMINA

Source	DF	Type I SS	Mean Square	F Value	Pr > F
VARIETE	2	93.3333333	46.6666667	3.50	0.0751
SOL	1	83.9007092	83.9007092	6.29	0.0334
SOL*VARIETE	2	222.7659574	111.3829787	8.35	0.0089

Source	DF	Type III SS	Mean Square	F Value	Pr > F
VARIETE	2	192.1276596	96.0638298	7.20	0.0135
SOL	1	123.7714286	123.7714286	9.28	0.0139
SOL*VARIETE	2	222.7659574	111.3829787	8.35	0.0089

FPSTAT INRA 1997 IV-Analyse de Variance-38

Tableau "Facteurs"

Dependent Variable : GERMINA

Source	DF	Type I SS	Mean Square	F Value	Pr > F
VARIETE	2	93.3333333	46.6666667	3.50	0.0751
SOL	1	83.9007092	83.9007092	6.29	0.0334
SOL*VARIETE	2	222.7659574	111.3829787	8.35	0.0089

Source	DF	Type III SS	Mean Square	F Value	Pr > F
VARIETE	2	192.1276596	96.0638298	7.20	0.0135
SOL	1	123.7714286	123.7714286	9.28	0.0139
SOL*VARIETE	2	222.7659574	111.3829787	8.35	0.0089

		Type I	
VARIETE	J-1	$SCE_{M_0} - SCE_{M'_1}$	$\dfrac{\left(SCE_{M_0} - SCE_{M'_1}\right) / J-1}{SCE_{M_3} / N-IJ}$
SOL	I-1	$SCE_{M'_1} - SCE_{M_2}$	$\dfrac{\left(SCE_{M'_1} - SCE_{M_2}\right) / I-1}{SCE_{M_3} / N-IJ}$
SOL*VARIETE	IJ-1	$SCE_{M_2} - SCE_{M_3}$	$\dfrac{\left(SCE_{M_2} - SCE_{M_3}\right) / IJ-1}{SCE_{M_3} / N-IJ}$

On constate la même chose que précédemment. En plus, il faut insister sur le fait que la valeur de la somme des carrés de type I pour un facteur diffère selon l'ordre d'introduction des facteurs.

Pour les sommes de type III, elles sont indifférentes à l'ordre des facteurs testés dans chaque modèle.

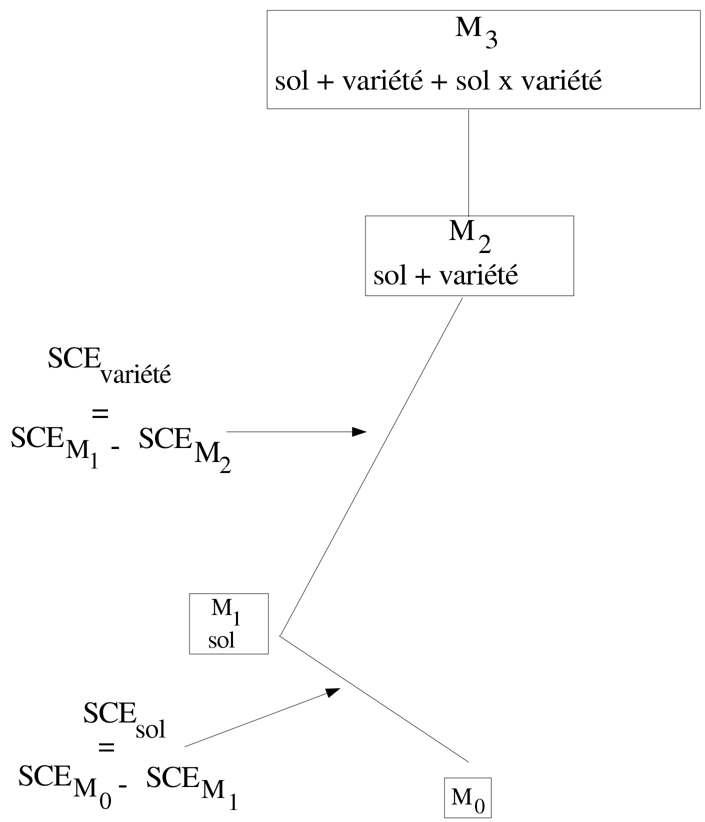
Plusieurs sommes de carrés sont disponibles.

Lesquelles choisir pour tester des modèles ?

Pourquoi ?

FPSTAT INRA 1997 IV-Analyse de Variance-40

SCE type I :



Ce premier transparent illustre le modèle

germination = sol + variété + sol x variété

SCE type I :

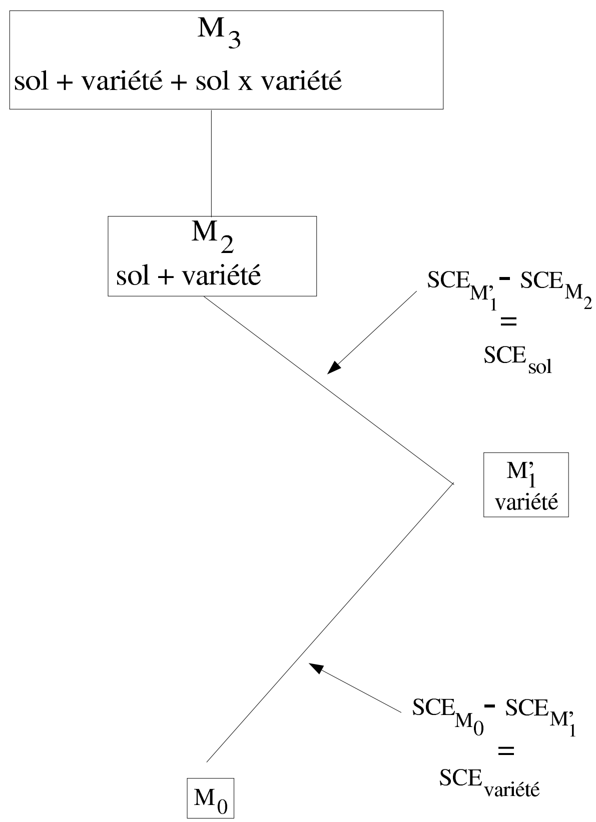


Illustration du modèle :

germination = variété + sol + sol x variété