Plan du Chapitre 4

Analyse de variance

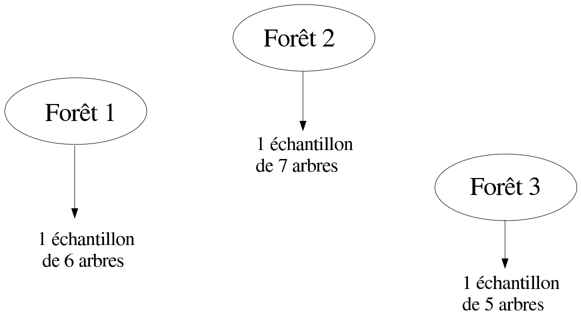
$\Updownarrow$

test d'une structuration des données

à l'aide de variables explicatives

de type qualitatif = facteurs

Analyse de variance à un facteur



Forêt 1	Forêt 2	Forêt 3
23.4	18.9	22.5
24.4	21.1	22.9
24.6	21.1	23.7
24.9	22.1	24.0
25.0	22.5	24.0
26.2	23.5	
	24.5	

Le vecteur des données

$$\mathrm{Y_n} = \begin{pmatrix} \mathrm{Y}_1 & = & 23.4 \\ \mathrm{Y}_2 & = & 24.4 \\ \mathrm{Y}_3 & = & 24.6 \\ \mathrm{Y}_4 & = & 24.9 \\ \mathrm{Y}_5 & = & 25.0 \\ \mathrm{Y}_6 & = & 26.2 \\ \mathrm{Y}_7 & = & 18.9 \\ \mathrm{Y}_8 & = & 21.1 \\ \mathrm{Y}_9 & = & 21.1 \\ \mathrm{Y}_{10} & = & 22.1 \\ \mathrm{Y}_{11} & = & 22.5 \\ \mathrm{Y}_{12} & = & 23.5 \\ \mathrm{Y}_{13} & = & 24.5 \\ \mathrm{Y}_{14} & = & 22.5 \\ \mathrm{Y}_{15} & = & 22.9 \\ \mathrm{Y}_{16} & = & 23.7 \\ \mathrm{Y}_{17} & = & 24.0 \\ \mathrm{Y}_{18} & = & 24.0 \end{pmatrix} \rightarrow \mathrm{Y_{ir}} = \begin{pmatrix} \mathrm{Y}_{11} & = & 23.4 \\ \mathrm{Y}_{12} & = & 24.4 \\ \mathrm{Y}_{13} & = & 24.6 \\ \mathrm{Y}_{14} & = & 24.9 \\ \mathrm{Y}_{15} & = & 25.0 \\ \mathrm{Y}_{16} & = & 26.2 \\ \mathrm{Y}_{21} & = & 18.9 \\ \mathrm{Y}_{22} & = & 21.1 \\ \mathrm{Y}_{23} & = & 21.1 \\ \mathrm{Y}_{24} & = & 22.1 \\ \mathrm{Y}_{25} & = & 22.5 \\ \mathrm{Y}_{26} & = & 23.5 \\ \mathrm{Y}_{27} & = & 24.5 \\ \mathrm{Y}_{31} & = & 22.5 \\ \mathrm{Y}_{32} & = & 22.9 \\ \mathrm{Y}_{33} & = & 23.7 \\ \mathrm{Y}_{34} & = & 24.0 \\ \mathrm{Y}_{35} & = & 24.0 \end{pmatrix}$$

La structuration en forêts explique t-elle
la variabilité des données ?

$\Updownarrow$

Les forêts sont elles différentes ?

Si oui : quelles forêts sont différentes ?

Modèle	SCE	Ecriture du modèle
M_1 I paramètres	SCE_{M_1}	$\mu_n (= \mu_{ir}) = \mu_i = \mu + \alpha_i$

$\downarrow$

Modèle	SCE	Ecriture du modèle
M_0 1 paramètre	SCE_{M_0}	$\mu_n (= \mu_{ir}) = \mu$

H_1

$\exists \mathrm{i}, \alpha_{\mathrm{i}} \neq 0$
$\exists (\mathrm{i}, \mathrm{i}') \; ; \mu_{\mathrm{i}} \neq \mu_{\mathrm{i}'}$

$\forall \mathrm{i}, \; \alpha_{\mathrm{i}} = 0$
$\forall (\mathrm{i}, \mathrm{i}'); \mu_{\mathrm{i}} = \mu_{\mathrm{i}'} = \mu$
$\Leftrightarrow \mu_1 = \mu_2 = \ldots \mu_{\mathrm{I}} = \mu$

H_0

Commandes et fichier sous SAS

```
data trv ;
infile 'foret3' ;
input foret hauteur ;
run ;
proc glm ;
class foret ;
model hauteur = foret ;
run ;
```

Fichier des données

1	23.4
1	24.4
1	24.6
1	24.9
1	25.0
1	26.2
2	18.9
2	21.1
2	21.1
2	22.1
2	22.5
2	23.5
2	24.5
3	22.5
3	22.9
3	23.7
3	24.0
3	24.0

General Linear Models Procedure

Class Level Information

Class	Levels	Values
Foret	3	1 2 3

Number of observations in data set = 18

General Linear Models Procedure

Dependent Variable : HAUTEUR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	25.30930159	12.65465079	7.30	0.0061
Error	15	26.00014286	1.73334286		
Corrected Total	17	51.30944444			

R-Square	C.V.	Root MSE	HAUTEUR Mean
0.493268	5.651840	1.316565	23.29444

General Linear Models Procedure

Dependent Variable : HAUTEUR

Source	DF	Type I SS	Mean Square	F Value	Pr > F
FORET	2	25.30930159	12.65465079	7.30	0.0061

Source	DF	Type III SS	Mean Square	F Value	Pr > F
FORET	2	25.30930159	12.65465079	7.30	0.0061

Tableau "Modèle"

General Linear Models Procedure

Class Level Information

Class	Levels	Values
Foret	3	1 2 3

Number of observations in data set = 18

General Linear Models Procedure

Dependent Variable : HAUTEUR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	25.30930159	12.65465079	7.30	0.0061
Error	15	26.00014286	1.73334286		
Corrected Total	17	51.30944444			

R-Square	C.V.	Root MSE	HAUTEUR Mean
0.493268	5.651840	1.316565	23.29444

Model	I-1	$SCE_{M_0} - SCE_{M_1}$	$\frac{SCE_{M_0} - SCE_{M_1}}{I-1}$	$\frac{\left(SCE_{M_0} - SCE_{M_1}\right)/ I-1}{SCE_{M_1}/ N-I}$
Error	N-I	SCE_{M_1}	$\frac{SCE_{M_1}}{N-I}$	
Corrected Total	N-1	SCE_{M_0}		

R-Square	C.V.	Root MSE	HAUTEUR Mean
$\frac{SCE_{M_0} - SCE_{M_1}}{SCE_{M_0}}$	100x(Root MSE)/Hauteur Mean	$\sqrt{\frac{SCE_{M_1}}{N-I}}$	Y..

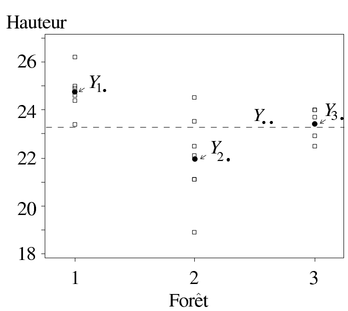
General Linear Models Procedure

Dependent Variable : HAUTEUR

Source	DF	Type I SS	Mean Square	F Value	Pr > F
FORET	2	25.30930159	12.65465079	7.30	0.0061

Source	DF	Type III SS	Mean Square	F Value	Pr > F
FORET	2	25.30930159	12.65465079	7.30	0.0061

FORET	I-1	$SCE_{M_0} - SCE_{M_1}$	$\dfrac{\left(SCE_{M_0} - SCE_{M_1}\right) / \text{I-1}}{SCE_{M_1} / \text{N-I}}$



$$SCE_{totale} = SCE_{facteur} + SCE_{résiduelle}$$

$$SCE_{totale} = \sum_n (Y_n - Y_{..})^2$$

$$SCE_{facteur} = \sum_i \sum_r (Y_{i.} - Y_{..})^2 = \text{variabilité inter} - \text{classes}$$

$$SCE_{résiduelle} = \sum_i \sum_r (Y_{ir} - Y_{i.})^2 = \text{variabilité intra} - \text{classe}$$

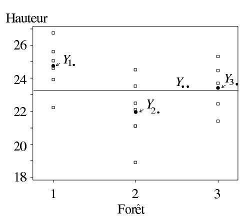
$$\mathrm{SCE_{totale}} = \mathrm{SCE_{M_0}} : \text{variabilité totale}$$

$$\mathrm{SCE_{facteur}} = \mathrm{SCE_{M_0}} - \mathrm{SCE_{M_1}} : \text{variabilité inter} - \text{classe}$$

$$\mathrm{SCE_{résiduelle}} = \mathrm{SCE_{M_1}} : \text{variabilité intra} - \text{classe}$$

d'où

$$\mathrm{F} = \frac{\mathrm{SCE_{facteur}} /\ \mathrm{I} - 1}{\mathrm{SCE_{résiduelle}} /\ \mathrm{N} - \mathrm{I}} = \frac{(\mathrm{SCE_{M_0}} - \mathrm{SCE_{M_1}}) /\ \mathrm{I} - 1}{\mathrm{SCE_{M_1}} /\ \mathrm{N} - \mathrm{I}}$$



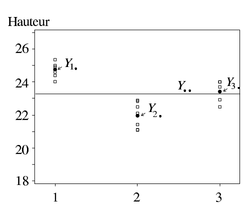
$\mathrm{Y_{1.}} = 24.75$

$\mathrm{Y_{2.}} = 21.96$

$\mathrm{Y_{3.}} = 23.42$

$(\mathrm{Y_{1.}} + \mathrm{Y_{2.}} + \mathrm{Y_{3.}})/3 = 23.38$

$\mathrm{H_0}$ ou $\mathrm{H_1}$?



$\mathrm{Y_{1.}} = 24.75$

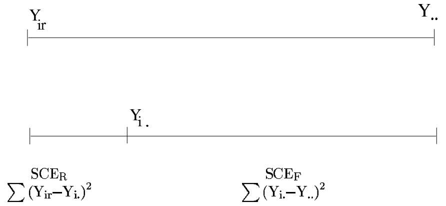
$\mathrm{Y_{2.}} = 21.96$

$\mathrm{Y_{3.}} = 23.42$

$(\mathrm{Y_{1.}} + \mathrm{Y_{2.}} + \mathrm{Y_{3.}})/3 = 23.38$

$\mathrm{H_0}$ ou $\mathrm{H_1}$?

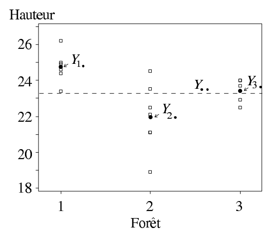
$$SCE_T : \sum (Y_{ir} - Y_{..})^2$$



$$\boxed{SCE_T = SCE_F + SCE_R}$$

si $SCE_F >> SCE_R \Rightarrow H_1$

si $SCE_F \simeq SCE_R \Rightarrow H_0$



Analyse de Variance à un facteur
Synthèse

TEST F modèle / ss modèle

M_1/M_0

Non Significatif

$\Rightarrow$ La structuration du Modèle M_1 est sans intérêt

$\Rightarrow$ On n'a pas mis en évidence de différences entre les niveaux du facteur

Significatif

$\Rightarrow$ La structuration du Modèle M_1 est acceptable

$\Rightarrow$ IL Y A des différences au risque α

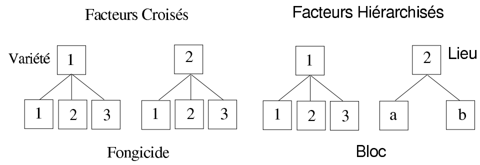
Lesquelles ?

Comparaison multiple de moyennes

Analyse de Variance à deux facteurs

2 cas

2 facteurs croisés | 1 facteur hiérarchisé à l'autre



Variété \ Fongicide	1	2	3
1	X	X	X
2	X	X	X

variété x lieu variété x fongicide	lieu / bloc bloc / sous-bloc
variété x testeur	portée / numéro lapin
concentration x température	famille génétique / numéro descendant
région x profession	région / département

Analyse de Variance à deux facteurs croisés

Facteur A, I niveaux

Facteur B, J niveaux

facteurs croisés

Variété \ Fongicide	1	2	3
1	X	X	X
2	X	X	X

les IJ combinaisons des niveaux des facteurs A et B peuvent être étudiées

2 cas

(dispositif) plan expérimental non orthogonal

(dispositif) plan expérimental orthogonal

Étude du nombre de jours avant germination pour des variétés de carottes

Y_{ijk}	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

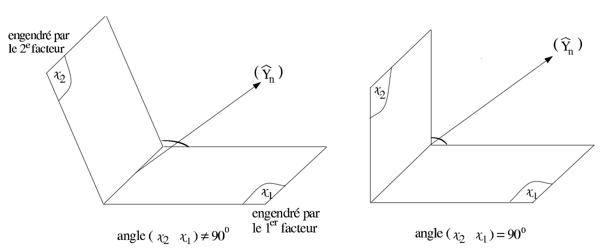
ex : dispositif déséquilibré

Étude de la teneur en huile de populations de tournesol

Origine	AFRIQUE	HONGRIE	MAROC
testeur 1	43.54 45.30	44.25 42.55	47.28 49.40
testeur 2	47.21 47.73	44.34 46.49	47.75 49.47

ex : dispositif équirépété

Illustration géométrique



Analyse de variance à deux facteurs croisés cas orthogonal

Exemple d'Applications de l'Analyse de la Variance dans le cas équirépété

Etude de la teneur en huile de populations de tournesol

Données expérimentales

Origine	AFRIQUE	HONGRIE	MAROC
testeur 1	43.54 45.30	44.25 42.55	47.28 49.40
testeur 2	47.21 47.73	44.34 46.49	47.75 49.47

Facteur A = testeur, I = 2 niveaux

Facteur B = origine, J = 3 niveaux

répétitions par combinaison AB, r = 2

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 (testeur)

$\mu_n = \mu_i = \mu + \alpha_i$

M'_1 (origine)

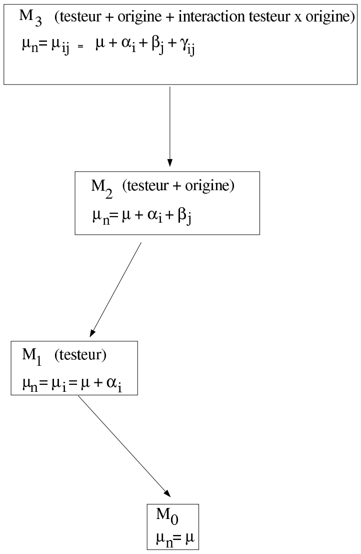
$\mu_n = \mu_j = \mu + \beta_j$

↘ ↙

M_0

$\mu_n = \mu$

1 1 43.54	data trv ;
1 1 45.30	infile 'tourn' ;
1 2 44.25	input testeur origine huile ;
1 2 42.55	run ;
1 3 47.28	proc glm ;
1 3 49.40	class testeur origine ;
2 1 47.21	model huile = testeur origine testeur*origine ;
2 1 47.73	run ;
2 2 44.34	
2 2 46.49	
2 3 47.75	
2 3 49.47	



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

General Linear Models Procedure

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

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