ANALYSE DE COVARIANCE ET GÉNÉRALISATION

Plan du chapitre 6

1. Analyse de Covariance “classique”

   L’intérêt d’introduire une covariable

2. Analyse de Covariance “classique”

   Comparaison de droites de régression parallèles

3. Première généralisation

   Comparaison de droites de régression quelconques

4. Seconde généralisation

   Niveaux d’un facteur pouvant être quantifiés

Objectif de ce chapitre

Dans le cadre du modèle linéaire général,

— Présenter ce que l’on appelle jusqu’à présent, de manière très restrictive, l’analyse de covariance et son intérêt.

— Généraliser l’analyse de covariance, aux modèles comportant des facteurs et des covariables.

— Présenter l’intérêt d’introduire des variables indicatrices.

— Montrer comment l’on peut passer progressivement à une quantification d’un facteur influençant une réponse donnée.

Analyse de covariance “classique” L’intérêt d’introduire une covariable

<u>Exemple : L’affinages des huîtres</u>

On teste l’effet de la position des paniers d’huîtres sur leur croissance.
Une pesée des paniers est réalisée en début d’expérimentation (INI) puis 1 mois après (FIN).

<u>La question</u>
L’effet de la position des paniers sur le poids en fin d’expérience peut-il être améliorer par la prise en compte du poids initial des paniers ?

Les données

obs	T	r	INI	FIN
1	1	1	27.2	32.6
2	1	2	32	36.6
3	1	3	33	37.7
4	1	4	26.8	31
5	2	1	28.6	33.8
6	2	2	26.8	31.7
7	2	3	26.5	30.7
8	2	4	26.8	30.4
9	3	1	28.6	35.2
10	3	2	22.4	29.1
11	3	3	23.2	28.9
12	3	4	24.4	30.2
13	4	1	29.3	35
14	4	2	21.8	27
15	4	3	30.3	36.4
16	4	4	24.3	30.5
17	5	1	20.4	24.6
18	5	2	19.6	23.4
19	5	3	25.1	30.3
20	5	4	18.1	21.8

Les modalités du traitement d’affinage sont :

1. près de l’arrivée d’eau de mer et au fond du bassin d’affinage
2. près de l’arrivée d’eau de mer et en surface
3. près de l’évacuation de l’eau du bassin et au fond
4. près de l’évacuation de l’eau du bassin et en surface
5. témoin dans la baie.

4 paniers de 10 huîtres sont placés aléatoirement dans chaque traitement. Il y a donc 4 répétitions (r). Chaque panier constitue l’unité expérimentale.

Commandes SAS

Modèle d'analyse de variance (orthogonale)

```
proc anova ;
class T ;
model FIN = T ;
run ;
```

Modèle linéaire général, modèle avec covariable

```
proc glm ;
class T ;
model FIN = INI T ;
run ;
```


FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-3

Sortie SAS

Modèle d'analyse de variance

Analysis of variance Procedure

Class Level Information

Class	Levels	Values
T	5	1 2 3 4 5

Number of observations in data set = 20

Dependent Variable : FIN

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	198.407000	49.601750	4.64	0.0122
Error	15	160.262500	**10.684167**		
Corrected Total	19	358.669500			

R-Square	C.V.	Root MSE	FIN Mean
0.553175	10.59706	3.26866	30.8450

Source	DF	Anova SS	Mean Square	F Value	Pr > F
T	4	**198.407000**	49.601750	**4.64**	0.0122

FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-4

Sortie SAS

Modèle avec covariable

General Linear Models Procedure

Class Level Information

Class	Levels	Values
T	5	1 2 3 4 5

Number of observations in data set = 20

General Linear Models Procedure

Dependent Variable : FIN

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	354.447177	70.889435	235.05	0.0001
Error	14	4.222323	**0.301595**		
Corrected Total	19	358.669500			

R-Square	C.V.	Root MSE	FIN Mean
0.988228	1.780438	0.54918	30.8450

Source	DF	Type I SS	Mean Square	F Value	Pr > F
INI	1	342.357817	342.357817	1135.16	0.0001
T	4	12.089359	3.022340	10.02	0.0005

Source	DF	Type III SS	Mean Square	F Value	Pr > F
INI	1	156.040177	156.040177	517.38	0.0001
T	4	**12.089359**	3.022340	**10.02**	0.0005

Les valeurs figurées en gras permettent de comparer les deux analyses.

On voit que de l'analyse de variance à l'analyse avec covariable :

— Le carré moyen des écarts résiduel (ligne “error”) passe de 10,684 à 0,30

— La somme des carrés des écarts traitement (ligne “T”) passe de 198, 407 à 12,089

— La statistique F passe de 4,64 à 10,02.

Ainsi la puissance du test augmente quand la covariable est incluse, car une bonne part de l'erreur dans l'ANOVA est due à des variations des valeurs initiales.

Analyse de covariance "classique"
Comparaison de droites de régression

Exemple : Pression sanguine des hommes et des femmes

On mesure la pression sanguine de personnes de différents âges.

La question

La variation de la pression sanguine (PS) en fonction de l'âge (AGE) est-elle différente selon que l'on mesure des hommes ou des femmes (SEXE) ?

Les données

Obs	SEXE	PS	AGE
1	Homme	158	41
2	Homme	185	60
3	Homme	152	41
4	Homme	159	47
5	Homme	176	66
6	Homme	156	47
7	Homme	184	68
8	Homme	138	43
9	Homme	172	68
10	Homme	168	57
11	Homme	176	65
12	Homme	164	57
13	Homme	154	61
14	Homme	124	36
15	Homme	142	44
16	Homme	144	50
17	Homme	149	47
18	Homme	128	19
19	Homme	130	22
20	Homme	138	21
21	Homme	150	38
22	Homme	156	52
23	Homme	134	41
24	Homme	134	18
25	Homme	174	51
26	Homme	174	55
27	Homme	158	65
28	Homme	144	33
29	Homme	139	23
30	Homme	180	70
31	Homme	165	56
32	Homme	172	62
33	Homme	160	51

On suppose, dans cet exemple, que la pente de la droite de la pression sanguine en fonction de l'âge est la même pour les deux sexes.

Obs	SEXE	PS	AGE
34	Homme	157	48
35	Homme	170	59
36	Homme	153	40
37	Homme	148	35
38	Homme	140	33
39	Homme	132	26
40	Homme	169	61
41	Femme	144	39
42	Femme	138	45
43	Femme	145	47
44	Femme	162	65
45	Femme	142	46
46	Femme	170	67
47	Femme	124	42
48	Femme	158	67
49	Femme	154	56
50	Femme	162	64
51	Femme	150	56
52	Femme	140	59
53	Femme	110	34
54	Femme	128	42
55	Femme	130	48
56	Femme	135	45
57	Femme	114	17
58	Femme	116	20
59	Femme	124	19
60	Femme	136	36
61	Femme	142	50
62	Femme	120	39
63	Femme	120	21
64	Femme	160	44
65	Femme	158	53
66	Femme	144	63
67	Femme	130	29
68	Femme	127	25
69	Femme	175	69

FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-6

Modèles proposés

Modèle de covariance

$$\mu_n = \theta_0 + \theta_1 \, \mathrm{age}_n + \underset{\substack{\downarrow \\ \text{facteur sexe}}}{\alpha_i} \qquad i = \text{Homme, Femme}$$

Modèle de régression avec variable indicatrice

Création d'une variable indicatrice dum

si SEXE = 'Homme' dum = 1

si SEXE = 'Femme' dum = 0

$$\mu_n = \theta_0^{'} + \theta_1^{'} \, \mathrm{age}_n + \theta_2 \, \mathrm{dum}_n$$

Les deux modèles permettent de modèliser une différence des ordonnées à l'origine.

Remarque :
Le modèle de régression avec variable indicatrice est une écriture irréductible du modèle de covariance ce qui est démontré dans le transparent suivant.

Si le facteur a I niveaux, I-1 variables indicatrices sont nécessaires pour écrire le modèle sous une forme irréductible.

Exemple : pour I = 3

si	T = 1	dum1 = 1
si	T = 2	dum1 = 0
si	T = 3	dum1 = 0
si	T = 1	dum2 = 0
si	T = 2	dum2 = 1
si	T = 3	dum2 = 0

Démonstration

Modèle de covariance

$$\begin{cases} \mu_{\mathrm{Fr}} = \theta_0 + \theta_1 \mathrm{age}_{\mathrm{Fr}} + \alpha_{\mathrm{F}} \\ \mu_{\mathrm{Hr}} = \theta_0 + \theta_1 \mathrm{age}_{\mathrm{Hr}} + \alpha_{\mathrm{H}} \end{cases}$$

Modèle de régression avec variable **indicatrice**

$$\begin{cases} \mu_{\mathrm{Fr}} = & \theta_0^{'} + \theta_1^{'} \mathrm{age}_{\mathrm{Fr}} \\ \mu_{\mathrm{Hr}} = & \theta_0^{'} + \theta_1^{'} \mathrm{age}_{\mathrm{Hr}} + \theta_2 \end{cases}$$

$\Leftrightarrow$ $\theta_0^{'} = \theta_0 + \alpha_{\mathrm{F}}$ ordonnée à l'origine pour les femmes

$\theta_1^{'} = \theta_1$ pente de la droite

$\theta_2 = \alpha_{\mathrm{H}} - \alpha_{\mathrm{F}}$ différence des ordonnées à l'origine

Rappel : Dans un modèle écrit sous une forme irréductible, les paramètres du modèle sont des **fonctions estimables**

Les paramètres du modèle de régression avec variable indicatrice sont d'interprétation plus facile, car ce modèle est sous une forme irréductible.

FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-8

Commande SAS

Modèle de covariance ;

```
proc glm ;
class sexe ;
model PS = age sexe ;
run ;
```

Modèle de régression avec variable indicatrice ;

```
if sexe ='Homme' then dum = 1 ;
if sexe ='Femme' then dum = 0 ;
proc reg ;
model PS = age dum ;
run ;
```


FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-9

Sortie SAS

Modèle de covariance

General Linear Models Procedure

Class Level Information

Class	Levels	Values
SEXE	2	Femme Homme

Number of observations in data set = 69

General Linear Models Procedure

Dependent Variable : PS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	17897.7652	8948.8826	113.08	0.0001
Error	66	5223.0463	79.1371		
Corrected Total	68	23120.8116			

R-Square	C.V.	Root MSE	PS Mean
0.774098	5.980292	8.89590	148.754

Source	DF	Type I SS	Mean Square	F Value	Pr > F
AGE	1	14868.2007	14868.2007	187.88	0.0001
SEXE	1	3029.5645	3029.5645	38.28	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
AGE	1	14003.9157	14003.9157	176.96	0.0001
SEXE	1	3029.5645	3029.5645	38.28	0.0001

On observe un effet âge et un effet sexe.

Donc l'ordonnée à l'origine est différente selon le sexe.

Sortie SAS

Modèle de régression avec variable indicatrice

Analysis of Variance

Dependent Variable : PS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	17897.76525	8948.88262	113.081	0.0001
Error	66	5223.04635	79.13707		
Corrected Total	68	23120.81159			

R-Square	C.V.	Root MSE	Dep Mean	Adj R-Sq
0.7741	5.98029	8.89590	148.75362	0.7673

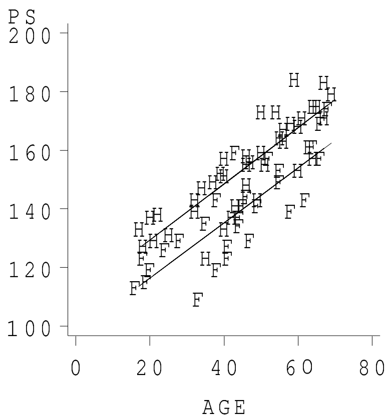
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0 : Parameter=0	Prob > |T|
INTERCEP	1	96.959922	3.62817487	26.724	0.0001
AGE	1	0.953452	0.07167443	13.303	0.0001
DUM	1	13.449325	2.17370376	6.187	0.0001

Les estimations des paramètres permettent aisément de visualiser les droites de régression.

FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-11

Visualisation graphique



FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-12

Première généralisation : comparaison de droites de régression quelconques

Exemple : Tarifs de cubage des sapins-épicéas

On mesure le **volume de la grume** .
On souhaite réaliser un **tarif de cubage** (= ajustement entre ce volume et une variable que l'on peut mesurer facilement sur arbre debout).
Cette variable est la **surface terrière** .
La question
Doit-on faire un seul tarif pour le sapin-épicéa ou bien un tarif pour chacune de ces deux essences ?

Les données

Obs	Espece	Volume	ST
1	Sapin	3	0.5
2	Sapin	2.5	2
3	Sapin	3.5	3
4	Sapin	3.5	3.5
5	Sapin	4	5.5
6	Sapin	4	8
7	Sapin	5	9
8	Sapin	5	11
9	Sapin	4.5	12.5
10	Sapin	6	13.5
11	Epicea	1.5	1
12	Epicea	2.5	2.5
13	Epicea	4	4
14	Epicea	3.5	5
15	Epicea	5	7
16	Epicea	6	7.5
17	Epicea	7.5	9.5
18	Epicea	8	11
19	Epicea	9	12.5
20	Epicea	10	13.5

grume=tronc commercialisé

$ST = \pi.D^2/4$ avec D^2 le diamètre du tronc mesuré 1,30 m du sol

Modèles proposés

Modèle de covariance avec **interaction** facteur-covariable

$$\mu_{\mathrm{n}} = \theta_0 + \theta_1\, \mathrm{ST_n} + \underbrace{\alpha_{\mathrm{i}}}_{\substack{\text{facteur}\\ \text{espèce}}} + \underbrace{\beta_{\mathrm{i}}\, .\, \mathrm{ST_n}}_{\substack{\text{interaction}\\ \text{facteur.covariable}}} \qquad \mathrm{i = Sapin,\ Epicéa}$$

Modèle de régression avec variables indicatrice et **auxiliaire**

Création d'une variable indicatrice

Si Espece = 'Sapin' dum = 1

Si Espece = 'Epicea' dum = 0

Création d'une variable auxiliaire

aux = dum * ST

$$\mu_{\mathrm{n}} = \theta_0^{'} + \theta_1^{'}\, \mathrm{ST_n} + \theta_2\, \mathrm{dum} + \theta_3\, \mathrm{aux}$$

Ces deux modèles permettent de modèliser

— une différence des ordonnées à l'origine

— une différence des pentes de droite de régression

Remarque :

Le modèle de régression avec variable indicatrice et auxiliaire est une écriture irréductible du modèle de covariance avec interaction facteur-covariable.

Ce qui est démontré dans le transparent suivant.

Si le facteur a I niveaux, I - 1 variables indicatrices et I - 1 variables auxiliaires sont nécessaires pour écrire le modèle sous une forme irréductible.

Démonstration

Modèle de covariance avec **interaction** facteur **covariable**

$$\mu_{\mathrm{Er}} = \theta_0 + \theta_1\ \mathrm{ST}_{\mathrm{Er}} + \alpha_{\mathrm{E}} + \beta_{\mathrm{E}}.\mathrm{ST}_{\mathrm{Er}}$$

$$\mu_{\mathrm{Sr}} = \theta_0 + \theta_1\ \mathrm{ST}_{\mathrm{Sr}} + \alpha_{\mathrm{S}} + \beta_{\mathrm{S}}.\mathrm{ST}_{\mathrm{Sr}}$$

Modèle de régression avec variables indicatrice et **auxiliaire**

$$\mu_{\mathrm{Er}} = \theta_0^{'} + \theta_1^{'}\ \mathrm{ST}_{\mathrm{Er}}$$

$$\mu_{\mathrm{Sr}} = \theta_0^{'} + \theta_1^{'}\ \mathrm{ST}_{\mathrm{Sr}} + \theta_2 + \theta_3\ \mathrm{ST}_{\mathrm{Sr}}$$

$\Leftrightarrow$ $\theta_0^{'} = \theta_0 + \alpha_{\mathrm{E}}$ ordonnée à l'origine pour l'Epicéa

$\theta_1^{'} = \theta_1 + \beta_{\mathrm{E}}$ pente de la droite pour l'Epicéa

$\theta_2 = \alpha_{\mathrm{S}} - \alpha_{\mathrm{E}}$ différence des ordonnées à l'origine

$\theta_3 = \beta_{\mathrm{S}} - \beta_{\mathrm{E}}$ différence des pentes

Les paramètres du modèle de régression avec variables indicatrice et auxiliaire sont d'interprètation plus facile car ce modèle est sous une forme irréductible

Commandes SAS

Modèle de covariance avec interaction facteur covariable

```
proc glm ;
class Espece ;
model Volume = ST Espece ST*Espece ;
run ;
```

Modèle de régression avec Variable indicatrice et auxiliaire

```
if Espece='Sapin' then dum=1 ;
if Espece='Epicea' then dum=0 ;
aux=ST*dum ;
proc reg ;
model Volume = ST dum aux ;
run ;
```


FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-16

Sortie SAS

Modèle de covariance avec interaction facteur-covariable

General Linear Models Procedure

Class Level Information

Class	Levels	Values
ESP	2	Epicea Sapin

Number of observations in data set = 20

General Linear Models Procedure

Dependent Variable : VOLUME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	93.9751998	31.3250666	177.43	0.0001
Error	16	2.8248002	0.1765500		
Corrected Total	19	96.8000000			

R-Square	C.V.	Root MSE	VOLUME Mean
0.970818	8.575072	0.42018	4.90000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ST	1	65.6270476	65.6270476	371.72	0.0001
ESPECE	1	9.6265324	9.6265324	54.53	0.0001
ST*ESPECE	1	18.7216198	18.7216198	106.04	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ST	1	67.5715845	67.5715845	382.73	0.0001
ESPECE	1	4.6201972	4.6201972	26.17	0.0001
ST*ESPECE	1	18.7216198	18.7216198	106.04	0.0001

Les trois effets sont significatifs, donc ordonnées à l'origine et pentes sont différentes.

Sortie SAS

Modèle de régression avec variable indicatrice et auxiliaire

Analysis of Variance

Dependent Variable : VOLUME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	93.97520	31.32507	177.429	0.0001
Error	16	2.82480	0.17655		
Corrected Total	19	96.80000			

R-Square	C.V.	Root MSE	Dep Mean	Adj R-Sq
0.9708	8.57507	0.42018	4.90000	0.9653

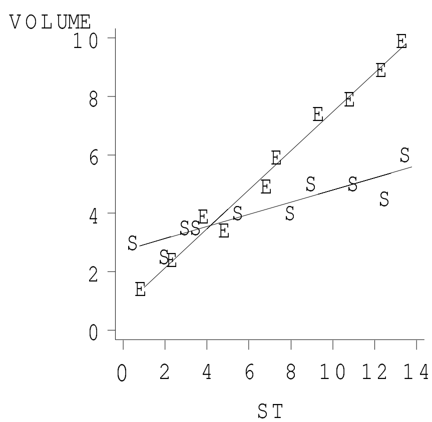
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0 : Parameter=0	Prob > |T|
INTERCEP	1	0.783469	0.27596985	2.839	0.0001
ST	1	0.668916	0.03290840	20.327	0.0001
DUM	1	1.894715	0.37037993	5.116	0.0001
AUX	1	-0.461351	0.04480165	-10.298	0.0001

Les estimations des paramètres permettent aisément de visualiser les droites de régression

FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-18

Visualisation graphique

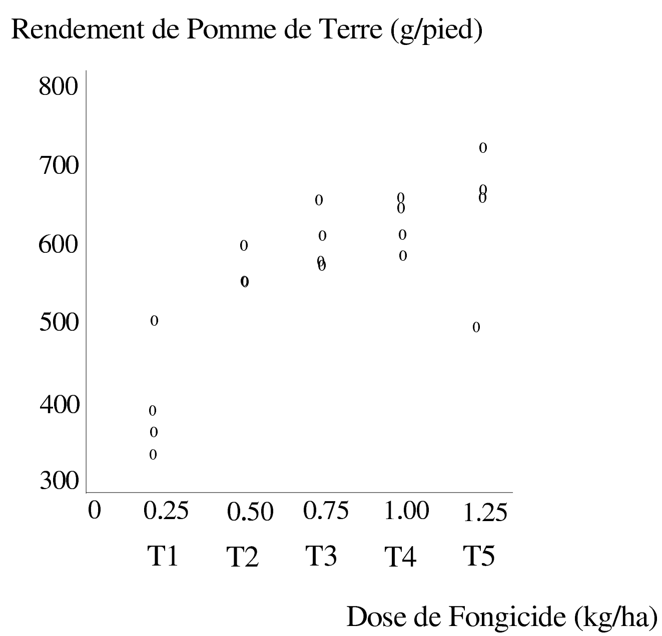


FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-19

Seconde généralisation : niveaux d'un facteur pouvant être quantifiés

<u>Exemple : Étude du rendement de pomme de terre en fonction du traitement (dose) de fongique</u>

Les données



T1 : 0,25kg/ha

T2 : 0,5

T3 : 0,75

T4 : 1,00

T5 : 1,25

4 répétitions

Modèle 1 : analyse de variance à un facteur (TRAIT)

$$\mu_{\mathrm{ir}} = \mu + \alpha_{\mathrm{i}}$$

i = 1... 5 traitements (facteur TRAIT)
r = 1... 4 répétitions
α_{i} effet du niveau i du facteur

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRAIT	4	133419.2	33354.8	8.81	0.0007
Erreur	15	56784.0	3785.6		
Totale	19	190203.2			

On veut déjà savoir s'il y a des différences entre les rendements obtenus pour les 5 traitements : ANALYSE DE VARIANCE.

Interprétation : le test F est hautement significatif. Les 5 traitements ne donnent pas des rendements équivalents.

Comment préciser ce résultat ?

Méthode de Newman-Keuls

$$\alpha_g = 5\%, \quad \widehat{\sigma}^2 = 3785.6, \quad \nu = 15\,(DF)$$

Nombre de moyennes	2	3	4	5
Etendue critique	92.7	113.0	125.4	134.3

Les moyennes portant la même lettre ne sont pas significativement différentes.

	Moyenne	TRAIT
A	629.0	T5
A	612.5	T4
A	600.5	T3
A	567.5	T2
B	404.5	T1

Par un test de comparaison de moyenne, par exemple la méthode de Newman-Keuls (équirépétitions).

On peut distinguer 2 groupes : T1 d'une part et T2 T3 T4 T5 d'autre part.

Ceci n'est pas très instructif, à part T1 qui est plus faible, les autres sont considérés comme équivalents.

FPSTAT INRA 1997 VI-Analyse de covariance et Généralisation-22

Modèle 2

$$Y_{ir} = \mu + \underbrace{\beta \times d_i + \alpha'_i}_{\alpha_i}$$

d_i	dose de fongicide correspondant au niveau i de TRAIT
α'_i	effet du facteur TRAIT, écart au modèle de régression linéaire

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DOSE = Régression linéaire	1	97614.4	97614.4	27.79	0.0001
TRAIT = Ecart à la régression	3	35804.8	11934.9	3.15	0.056
Erreur	15	56784.0	3785.6		
Totale	19	190203.2			

Le deuxième modèle envisagé décompose α_i en

— $\beta \times d_i$: régression linéaire simple en fonction de la dose

— $+\alpha'_i$: ce qui n'est pas expliqué par $\beta \times d_i$ (écart entre α_i et la régression linéaire).

Avec les sommes de carrés type I, on décompose la SCE du terme TRAIT du modèle 1 en SCE expliquée par la régression et SCE expliquée par l'effet TRAIT non pris en compte par la régression.

L'effet linéaire DOSE est hautement significatif.

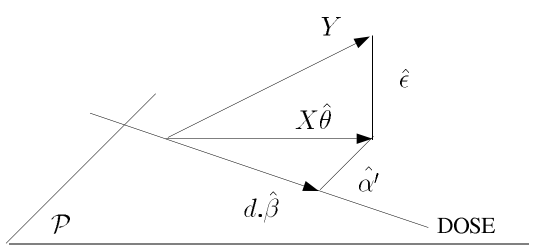
L'effet TRAIT est pratiquement significatif à 5 %.

Remarque :

Pour le terme TRAIT la SCE de type I est égale à la SCE de type II, car c'est le dernier effet introduit dans le modèle. C'est bien ce que l'on cherche à regarder :

ce qui est expliqué par TRAIT **après** la régression sur DOSE

Par contre, pour l'effet DOSE, comme il est contenu dans l'espace engendré par TRAIT, son pouvoir explicatif après TRAIT est nul. Il faut donc bien regarder son pouvoir explicatif **avant** TRAIT.



$\mathcal{P}$ engendré par TRAIT

$$
\begin{aligned}
\mathrm{Y} &= \mathrm{X}\widehat{\theta} + \widehat{\varepsilon} && \text{projection 1 sur } \mathcal{P} \\
\mathrm{X}\widehat{\theta} &= d.\widehat{\beta} + \widehat{\alpha'} && \text{projection 2 sur DOSE}
\end{aligned}
$$

$$
\begin{pmatrix} \alpha_1 \\ \alpha_1 \\ \alpha_1 \\ \alpha_1 \\ \vdots \\ \alpha_5 \\ \alpha_5 \\ \alpha_5 \\ \alpha_5 \end{pmatrix} = \begin{pmatrix} \mathrm{d}_1 \\ \mathrm{d}_1 \\ \mathrm{d}_1 \\ \mathrm{d}_1 \\ \vdots \\ \mathrm{d}_5 \\ \mathrm{d}_5 \\ \mathrm{d}_5 \\ \mathrm{d}_5 \end{pmatrix} .\beta + \begin{pmatrix} \alpha_1' \\ \alpha_1' \\ \alpha_1' \\ \alpha_1 \\ \vdots \\ \alpha_5' \\ \alpha_5' \\ \alpha_5' \\ \alpha_5 \end{pmatrix}
$$

Illustration géométrique :

— On projette Y sur l'espace $\mathcal{P}$ engendré par les traitements (TRAIT).

$\Rightarrow$on obtient $X\widehat{\theta}$.

— On projette $X\widehat{\theta}$ sur la droite engendrée par le vecteur DOSE.

$\Rightarrow d.\widehat{\beta}, \widehat{\alpha'}, \widehat{\epsilon}$ sont 2 à 2 orthogonaux donc leurs 3 sommes de carrés s'ajoutent.

Modèle 3

$$Y_{ir} = \mu + \underbrace{\beta \times d_i + \gamma \times d_i^2 + \alpha_i''}_{\alpha_i}$$

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DOSE	1	97614.4	97614.4	27.79	0.0001
DOSE^2	1	28170.3	28170.3	7.44	0.0156
TRAIT = Ecart au modèle quadratique	2	7634.5	3817.3	1.01	0.3883
Erreur	15	56784.0	3785.6		
Totale	19	190203.2			

On peut s'arrêter là et conclure que la relation entre le rendement et la dose peut être décrite par une régression linéaire simple mais cela semble grossier.

On peut donc envisager un troisième modèle qui décompose l'effet α_i en

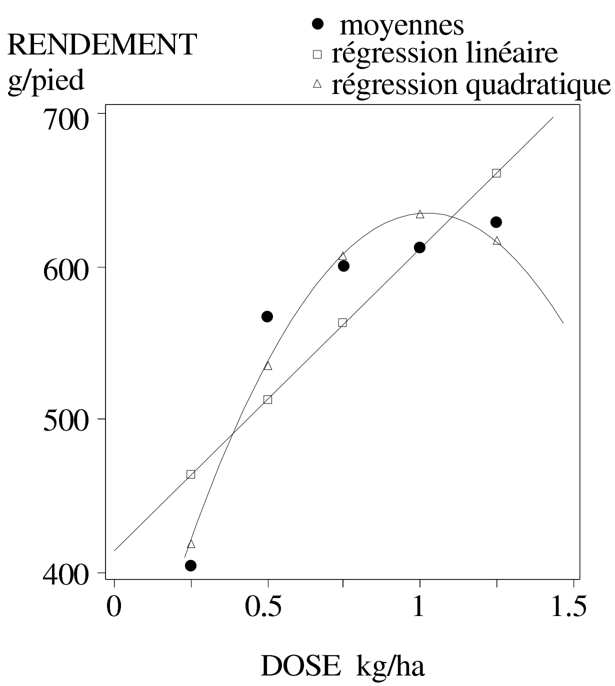
— $\beta \times d_i$: régression linéaire

— $+\gamma \times d_i^2$: terme où la valeur de dose est au carré (régression quadratique). Ceci permet de tester si la relation Rendement/Dose présente une courbure.

— $+\alpha_i''$: effet TRAIT non pris en compte ou écart au modèle quadratique.

L'effet quadratique DOSE^2 apporte une part d'explication significative.

L'effet TRAIT (écart au modèle quadratique) n'est plus du tout significatif, il est complétement expliqué par la régression DOSE + DOSE^2.



Le graphique présente les moyennes (●) du Rendement par niveau de Dose.

On y a superposé la droite de régression linéaire (□) qui s'ajuste au mieux et la parabole de régression quadratique (△) s'ajustant au mieux.

On comprend ainsi pourquoi le Modèle 3 précédemment vu ne présente plus du tout d'effet TRAIT significatif (écart à la régression quadratique non significatif).

En revanche le Modèle 2 présentait un effet TRAIT pratiquement significatif à 5 %, la droite ajustée ne décrit que très grossièrement la relation.

CONCLUSION : la parabole approxime de façon correcte la relation Rendement/Dose même si ce graphique indique que ce n'est peut être pas le modèle idéal.

Analyse de covariance et généralisation
Conclusions

L’analyse de covariance classique s’inscrit

comme la régression et l’analyse de variance dans l’ensemble plus général des modèles linéaires.

Replacée dans cet ensemble

elle peut être généralisée à bien d’autres situations.