

CHAPITRE 10

ANOVA à deux ou trois facteurs aléatoires et mixtes

Dans ce chapitre, on s'intéresse à l'analyse des plans qui font intervenir plusieurs facteurs où au moins un facteur est représenté par des modalités aléatoires choisies au hasard dans une population de modalités.

On parle d'un modèle à *effets aléatoires* lorsque tous les facteurs de l'expérience sont aléatoires et d'un modèle à *effet mixtes* lorsqu'on retrouve des facteurs fixes et des facteurs aléatoires. Nous ne considérerons que le cas d'un plan équilibré. Dans un premier nous étudions les modèles à deux facteurs équilibrés. Soit n la taille d'échantillon commune des IJ échantillons obtenus pour chaque traitement.

10.1 Modèle à deux facteurs aléatoires

On suppose que l'observation Y_{ijk} s'écrit:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, n,$$

où μ est une constante et α_i , β_j , γ_{ij} et e_{ijk} des variables aléatoires indépendantes telles que

$$\alpha_i \sim N(0, \sigma_\alpha^2)$$

$$\begin{aligned}\beta_j &\sim N(0, \sigma_\beta^2) \\ \gamma_{ij} &\sim N(0, \sigma_\gamma^2) \\ e_{ijk} &\sim N(0, \sigma^2)\end{aligned}$$

Sous ces hypothèses, on a $Y_{ijk} \sim N(\mu, \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma^2)$. Les modèles à effets aléatoires servent en pratique à faire l'étude poussée de la variabilité d'une réponse en identifiant les sources de cette variabilité.

Les différentes moyennes $\bar{Y}_{ij.}$, $\bar{Y}_{i..}$, $\bar{Y}_{.j.}$ et $\bar{Y}_{...}$ sont définies d'une manière analogue au chapitre précédent:

$$\begin{aligned}\bar{Y}_{ij.} &= \frac{1}{n} \sum_{k=1}^n Y_{ijk} \quad \text{moyenne d'une cellule} \\ \bar{Y}_{i..} &= \frac{1}{J} \sum_{j=1}^J \bar{Y}_{ij.} \quad \text{moyenne d'une ligne} \\ \bar{Y}_{.j.} &= \frac{1}{I} \sum_{i=1}^I \bar{Y}_{ij.} \quad \text{moyenne d'une colonne} \\ \bar{Y}_{...} &= \frac{1}{I} \sum_{i=1}^I \bar{Y}_{i..} \quad \text{moyenne de toutes les observations} \\ &= \frac{1}{J} \sum_{j=1}^J \bar{Y}_{.j.} \\ &= \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \bar{Y}_{ij.}\end{aligned}$$

Ces moyennes s'écrivent dans ce contexte:

$$\begin{aligned}\bar{Y}_{ij.} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \bar{e}_{ij.} \\ \bar{Y}_{i..} &= \mu + \alpha_i + \bar{\beta} + \bar{\gamma}_i + \bar{e}_{i..} \\ \bar{Y}_{.j.} &= \mu + \bar{\alpha} + \beta_j + \bar{\gamma}_j + \bar{e}_{.j.} \\ \bar{Y}_{...} &= \mu + \bar{\alpha} + \bar{\beta} + \bar{\gamma}_.. + \bar{e}_{...}\end{aligned}$$

On a vu au chapitre précédent que la somme total des carrés des erreurs $SST = \sum_{i,j,k} (Y_{ijk} -$

$\bar{Y}_{...}$)² s'écrit sous la forme

$$SST = SSA + SSB + SS(AB) + SSE = SS_{model} + SSE$$

où SSA , SSB , SST et $SS(AB)$ sont des variables aléatoires indépendantes définies par:

$$\begin{aligned} SSA &= nJ \sum_{i=1}^I \{\bar{Y}_{i..} - \bar{Y}_{...}\}^2 \\ SSB &= nI \sum_{j=1}^J \{\bar{Y}_{.j.} - \bar{Y}_{...}\}^2 \\ SS(AB) &= n \sum_{i=1}^I \sum_{j=1}^J \{\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\}^2 \\ SSE &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^n \{Y_{ijk} - \bar{Y}_{ij.}\}^2 \end{aligned}$$

Cette égalité reste valide dans ce contexte. Examinons de plus près chacune de ces 4 statistiques.

Pour la statistique SSA , on remarque que $\bar{Y}_{i..} - \bar{Y}_{...} = \{\alpha_i + \bar{\gamma}_i + \bar{e}_{i..}\} - \{\bar{\alpha} + \bar{\gamma}_{..} + \bar{e}_{...}\}$. On en déduit que SSA ne dépend pas de σ_β^2 . Et on démontre que

$$\frac{(I-1)MSA}{\sigma^2 + n\sigma_\gamma^2 + nJ\sigma_\alpha^2} \sim \chi_{I-1}^2$$

De même, on montre que

$$\frac{(J-1)MSB}{\sigma^2 + n\sigma_\gamma^2 + nI\sigma_\beta^2} \sim \chi_{J-1}^2$$

Pour SSE , on remarque que $Y_{ijk} - \bar{Y}_{ij.} = e_{ijk} - \bar{e}_{ij.}$. Cette statistique ne dépend donc que de σ^2 . On montre que

$$\frac{(n-1)IJMSE}{\sigma^2} \sim \chi_{IJ(n-1)}^2.$$

Enfin, pour $SS(AB)$, on remarque que

$$\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...} = \{\gamma_{ij} - \bar{\gamma}_i - \bar{\gamma}_j + \bar{\gamma}_{..}\} + \{\bar{e}_{ij.} - \bar{e}_{i..} - \bar{e}_{.j.} + \bar{e}_{...}\}$$

et on montre que

$$\frac{(I-1)(J-1)MS(AB)}{\sigma^2 + n\sigma_\gamma^2} \sim \chi_{(I-1)(J-1)}^2$$

Source de Variation	Statistique	Loi
Facteur A	$\frac{(I-1)MSA}{\sigma^2 + n\sigma_\gamma^2 + nI\sigma_\alpha^2}$	χ_{I-1}^2
Facteur B	$\frac{(J-1)MSB}{\sigma^2 + n\sigma_\gamma^2 + nI\sigma_\beta^2}$	χ_{J-1}^2
Interaction AB	$\frac{(I-1)(J-1)MS(AB)}{\sigma^2 + n\sigma_\gamma^2}$	$\chi_{(I-1)(J-1)}^2$
Erreur	$\frac{(n-1)JMSE}{\sigma^2}$	$\chi_{IJ(n-1)}^2$

On récapitule les résultats dans le tableau suivant:

Ce tableau suggère les tests suivant:

La région de rejet du test $H_0 : \sigma_\gamma^2 = 0$ versus $H_0 : \sigma_\gamma^2 > 0$ s'écrit:

$$\frac{MS(AB)}{MSE} > F_{(I-1)(J-1), IJ(n-1), \alpha}.$$

La région de rejet du test $H_0 : \sigma_\alpha^2 = 0$ versus $H_0 : \sigma_\alpha^2 > 0$ s'écrit:

$$\frac{MSA}{MS(AB)} > F_{I-1, (I-1)(J-1), \alpha}.$$

La région de rejet du test $H_0 : \sigma_\beta^2 = 0$ versus $H_0 : \sigma_\beta^2 > 0$ s'écrit:

$$\frac{MSB}{MS(AB)} > F_{J-1, (I-1)(J-1), \alpha}.$$

Ce tableau nous permet aussi de donner des expressions pour les fonctions de puissance de ces tests.

D'autre part, il est facile de voir que les paramètres μ et σ^2 sont estimés par $\bar{Y}_{..}$ et MSE respectivement. Les variances σ_α^2 , σ_β^2 et σ_γ^2 sont quand à elles estimées respectivement par

$$\begin{aligned} \hat{\sigma}_\alpha^2 &= \max\left(\frac{MSA - MS(AB)}{nJ}, 0\right) \\ \hat{\sigma}_\beta^2 &= \max\left(\frac{MSB - MS(AB)}{nI}, 0\right) \\ \hat{\sigma}_\gamma^2 &= \max\left(\frac{MS(AB) - MSE}{I}, 0\right) \end{aligned}$$

Les dérivations sont similaires au cas d'un modèle avec un seul facteur aléatoire.

10.2 Modèle mixte à deux facteurs

On utilise ce modèle pour analyser un schéma factoriel où les modalités d'un facteur sont fixes et ceux de l'autre sont aléatoires. Si A et B sont respectivement les facteurs fixes et aléatoires, le modèle s'écrit:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, n,$$

où μ et les α_i sont des constantes vérifiant $\sum_{i=1}^I \alpha_i = 0$, les β_j et les e_{ijk} sont des variables aléatoires indépendantes les unes des autres telles que

$$\begin{aligned} \beta_j &\sim N(0, \sigma_\beta^2) \\ e_{ijk} &\sim N(0, \sigma^2). \end{aligned}$$

Le cas des γ_{ij} est un peu délicat. Au fait il existe deux approches: modèles *restreint* et modèle *non restreint*.

10.2.1 Modèle mixte non-restreint

Dans le modèle mixte non-restreint, on suppose que les γ_{ij} sont indépendantes et proviennent d'une loi $N(0, \sigma_\gamma^2)$. En d'autres termes, on suppose que le terme d'interaction est un facteur à effet aléatoire. Cette solution est adoptée par la majorité des logiciels. Sous ce modèle, on a le tableau suivant:

Source de Variation	Statistique	Loi
Facteur A	$\frac{(I-1)MSA}{\sigma^2 + n\sigma_\gamma^2}$	$\chi_{I-1}^2 \left(\frac{Jn \sum_{i=1}^I \alpha_i^2}{\sigma^2 + n\sigma_\gamma^2} \right)$
Facteur B	$\frac{(J-1)MSB}{\sigma^2 + n\sigma_\gamma^2 + nI\sigma_\beta^2}$	χ_{J-1}^2
Interaction AB	$\frac{(I-1)(J-1)MS(AB)}{\sigma^2 + n\sigma_\gamma^2}$	$\chi_{(I-1)(J-1)}^2$
Erreur	$\frac{(n-1)IJMSE}{\sigma^2}$	$\chi_{IJ(n-1)}^2$

La grande différence par rapport au modèle à deux facteurs aléatoires se situe au niveau de *SSA*. En effet, sous ce modèle, on a

$$\bar{Y}_{i..} - \bar{Y}_{...} = \alpha_i + \{\bar{\gamma}_{i.} + \bar{e}_{i..}\} - \{\bar{\gamma}_{..} + \bar{e}_{...}\}.$$

Par conséquent, $E[\bar{Y}_{i..} - \bar{Y}_{...}] \neq 0$ et *MSA* suit une khi-deux non-centrée.

Les régions de rejet pour les trois tests du paragraphes précédent restent inchangées. Seule la fonction de puissance du test $H_0 : \alpha_1 = \dots = \alpha_I = 0$ change. L'étudiant doit être capable de l'écrire.

On estime μ par $\bar{Y}_{...}$, σ^2 par *MSE* et α_i par $\bar{Y}_{i..} - \bar{Y}_{...}$. Les estimateurs des variances $\hat{\sigma}_\beta^2$ et $\hat{\sigma}_\gamma^2$ demeurent aussi inchangés par rapport au modèle à deux facteurs aléatoires, c'est à dire:

$$\begin{aligned}\hat{\sigma}_\beta^2 &= \max\left(\frac{MSB - MS(AB)}{nI}, 0\right) \\ \hat{\sigma}_\gamma^2 &= \max\left(\frac{MS(AB) - MSE}{I}, 0\right)\end{aligned}$$

10.2.2 Modèle mixte restreint

Dans un modèle mixte restreint, pour chaque $j = 1, \dots, J$, on définit le vecteur aléatoire $\gamma_{.j}$ par

$$\gamma_{.j} = \begin{pmatrix} \gamma_{1j} \\ \gamma_{2j} \\ \vdots \\ \gamma_{Ij} \end{pmatrix}$$

On suppose que les $\gamma_{.j}$ sont indépendents et proviennent d'une loi normale multivariée de moyenne nulle et de matrice de variance-covariance égale à $\sigma_\gamma^2 \Sigma$ où:

$$\Sigma = \begin{pmatrix} 1 - \frac{1}{I} & -\frac{1}{I} & -\frac{1}{I} & \dots & -\frac{1}{I} \\ -\frac{1}{I} & 1 - \frac{1}{I} & -\frac{1}{I} & \dots & -\frac{1}{I} \\ -\frac{1}{I} & -\frac{1}{I} & 1 - \frac{1}{I} & \dots & -\frac{1}{I} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{I} & -\frac{1}{I} & -\frac{1}{I} & \dots & 1 - \frac{1}{I} \end{pmatrix}$$

Sous ces conditions, on montre que $Var[\sum_{i=1}^I \gamma_{ij}] = 0$ et par conséquent, $\sum_{i=1}^I \gamma_{ij} = 0$ pour tout $j = 1, \dots, J$. Le terme d'interaction γ_{ij} est donc un terme mixte.

Sous ce modèle, on a le tableau suivant:

Source de Variation	Statistique	Loi
Facteur A	$\frac{(I-1)MSA}{\sigma^2 + n\sigma_\gamma^2}$	$\chi_{I-1}^2 \left(\frac{Jn \sum_{i=1}^I \alpha_i^2}{\sigma^2 + n\sigma_\gamma^2} \right)$
Facteur B	$\frac{(J-1)MSB}{\sigma^2 + nI\sigma_\beta^2}$	χ_{J-1}^2
Interaction AB	$\frac{(I-1)(J-1)MS(AB)}{\sigma^2 + n\sigma_\gamma^2}$	$\chi_{(I-1)(J-1)}^2$
Erreur	$\frac{(n-1)IJMSE}{\sigma^2}$	$\chi_{IJ(n-1)}^2$

D'après ce tableau, la région de rejet du test $H_0 : \sigma_\beta^2 = 0$ versus $H_0 : \sigma_\beta^2 > 0$ devient alors:

$$\frac{MSB}{MSE} > F_{J-1, IJ(n-1), \alpha}.$$

Seule l'estimation de σ_β^2 change par rapport au modèle non restreint. En effet, d'après le tableau ci haut, on a:

$$\hat{\sigma}_\beta^2 = \max\left(\frac{MSB - MSE}{nI}, 0\right).$$

10.3 Discussion

Jusqu'à présent, ce chapitre traite d'expériences complètement randomisées, avec des facteurs *croisés*. On dit que des facteurs sont croisés si un traitement est une combinaison de modalités des facteurs à l'étude.

En pratique, les expériences avec des facteurs aléatoires croisés sont peu fréquentes. L'objectif de ces expérience est d'estimer les composantes de la variance associées aux facteurs et à leurs interactions. Dans un usine on peut vouloir estimer la variabilité du produit manufacturé associée à la variabilité entre les lots des fournisseurs A et B. Parfois un facteur qui a priori serait considéré comme étant fixe (par exemple la température) peut être défini

comme étant aléatoire si ses modalités sont un échantillon représentatif des conditions qui prévalent lors de la fabrication d'un certain produit.

Notons finalement qu'il y a une correspondance entre les modèles restreints et non restreints des sections précédentes. En effet, le modèle non restreint,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, n,$$

où μ et les α_i sont des constantes vérifiant $\sum_{i=1}^I \alpha_i = 0$ peut se réécrire

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \bar{\gamma}_{\cdot j} + \gamma_{ij} - \bar{\gamma}_{\cdot j} + e_{ijk}.$$

Ceci est un modèle restreint

$$Y_{ijk} = \mu + \alpha_i + \beta_j^r + \gamma_{ij}^r + e_{ijk}$$

avec $\beta_j^r = \beta_j + \bar{\gamma}_{\cdot j}$ et variance $\sigma_\beta^2 + \sigma_\gamma^2/I$ et $\gamma_{ij}^r = \gamma_{ij} - \bar{\gamma}_{\cdot j}$ et σ_γ^2 comme paramètre de variance.

10.4 Plans en parcelles partagées (schéma split plot)

Un schéma split plot est une généralisation d'un plan randomisé avec blocs. Pour introduire cette notion, il est utilisé d'utiliser une expérience en agriculture.

On veut étudier simultanément les types de PLANT ($I = 3$ modalités) et la CONCENTRATION d'engrais ($K = 3$ modalités) sur la production (Y) de fraise. Pour ce faire on dispose de $J = 3$ champs (ou blocs). Il y a en tout 9 traitements (PLANTS, CONCENTRATION) et on dispose de 3 blocs pour réaliser l'expérience.

Pour planifier cette expérience on pourrait utiliser un plan randomisé avec blocs. On diviserait chaque champ en 9 parcelles et on randomiserait les 9 traitements dans chaque bloc.

Supposons maintenant qu'un des deux facteurs, disons CONCENTRATION, est plus important et que l'on aimerait réduire au maximum l'erreur expérimentale pour ce facteur. On pourrait procéder en deux temps:

1. On divise chaque bloc en 3 grandes parcelles et on randomise les modalités de PLANT dans chaque grande parcelle. On dispose ainsi de 9 grandes parcelles et un schéma avec

blocs est utilisé pour comparer les modalités de PLANT (unité d'observation=grande parcelle).

2. Pour assigner la modalité de CONCENTRATION, on traite chaque grande parcelle comme un bloc. On la divise en 3 et on assigne au hasard une CONCENTRATION à chaque sous-parcelle (unité d'observation=sous-parcelle).

Avec ce type d'expérience il y a deux niveaux d'analyse. Pour le facteur PLANT les unités d'observation sont les grandes parcelles et on a un plan avec blocs ordinaire. Par contre, pour CONCEN, et l'interaction PLANT×CONCEN les unités sont les sous parcelles. Donc il y a deux parties à l'analyse; une est intra grande parcelles l'autre est inter grande parcelles. En général, l'erreur expérimentale est plus petite pour l'analyse intra grande parcelles. Le modèle pour les données s'écrit:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + \epsilon_{ijk},$$

avec

$i = 1, 2, 3$ représente la modalité de PLANT

$j = 1, 2, 3$ est associé aux blocs

$k = 1, 2, 3, 4$ représente la modalité de CONCEN.

Les facteurs aléatoires sont $\beta_j \sim N(0, \sigma_\beta^2)$, $(\alpha\beta)_{ij} \sim N(0, \sigma_{\alpha\beta}^2)$ et $\epsilon_{ijk} \sim N(0, \sigma^2)$. En général, on utilise un modèle non restreint qui n'impose aucune contrainte aux paramètres d'interaction. La table ANOVA pour ce modèle est donnée au tableau 10.1

À faire: Calculer les espérances des sommes de carrés moyens pour le modèle split plot.

Noter que la somme de carrés BLOC* PLANT est est la somme de carrés erreur pour l'analyse inter grandes parcelles. Notons que la somme de carrés des erreurs peut être décomposé en SSBLOC*CONCEN+ SSBLOC*PLANT*CONCEN. La première somme de carré peut être utilisé comme terme d'erreur pour CONCEN et la deuxième pour l'interaction PLANT*CONCEN.

Source	d.d.l.	Sommes de carrés	F
BLOC	$J - 1$	$KI \sum_j (\bar{Y}_{.j} - \bar{Y}_{...})^2$	$MS_{\text{BLOC}}/MS_{\text{B} * \text{P}}$
PLANT	$I - 1$	$KJ \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$	$MS_{\text{PLANTS}}/MS_{\text{B} * \text{P}}$
BLOC*PLANT	$(J - 1)(I - 1)$	$K \sum_{i,j} (\bar{Y}_{ij.} - \bar{Y}_{.j} - \bar{Y}_{i..} + \bar{Y}_{...})^2$	
CONCEN	$K - 1$	$IJ \sum_i (\bar{Y}_{i..K} - \bar{Y}_{...})^2$	MS_{CONCEN}/MSE
PLANT*CONCEN	$(K - 1)(I - 1)$	$J \sum_{i,k} (\bar{Y}_{i.k} - \bar{Y}_{i..} - \bar{Y}_{.k} + \bar{Y}_{...})^2$	$MS_{\text{P} * \text{C}}/MSE$
Erreur	$I(J - 1)(K - 1)$	$\sum_{i,j,k} (\bar{Y}_{ijk} - \bar{Y}_{i.k} - \bar{Y}_{ij.} + \bar{Y}_{i..})^2$	

TABLEAU 10.1: Table ANOVA pour un schéma split-plot

EXEMPLE 10.1 *Un fabricant de papier veut étudier l'impact de la PRÉPARATION de la pâte (3 modalités) et la TEMPÉRATURE de cuisson de la pâte (4 modalités) sur la force du papier produit. Il veut faire 3 répétitions des 12 traitements. Durant une journée de travail, l'usine peut seulement traiter 12 lots de pâte de papier. Trois JOURS sont nécessaires pour réaliser l'expérience; ce sont des blocs. A chaque jour il procède de la façon suivante. Il choisit un mode de préparation au hasard et fait un lot de pâte qu'il divise en 4 échantillons. Chaque échantillon est séché à une des température à l'étude. On est donc en présence d'un schéma split-plot où les blocs sont les jours et les lots de pâte sont les grandes parcelles.*

Le graphique d'interaction suggère la présence d'une faible interaction entre les deux facteurs. Le programme SAS pour analyser des données est:

```

data pate;
do bloc=1 to 3;
do method=1 to 3;
do te=1 to 4;
input force @;
temp=200+(te-1)*25; output; end; end; end; drop te;
datalines;
30 35 37 36

```

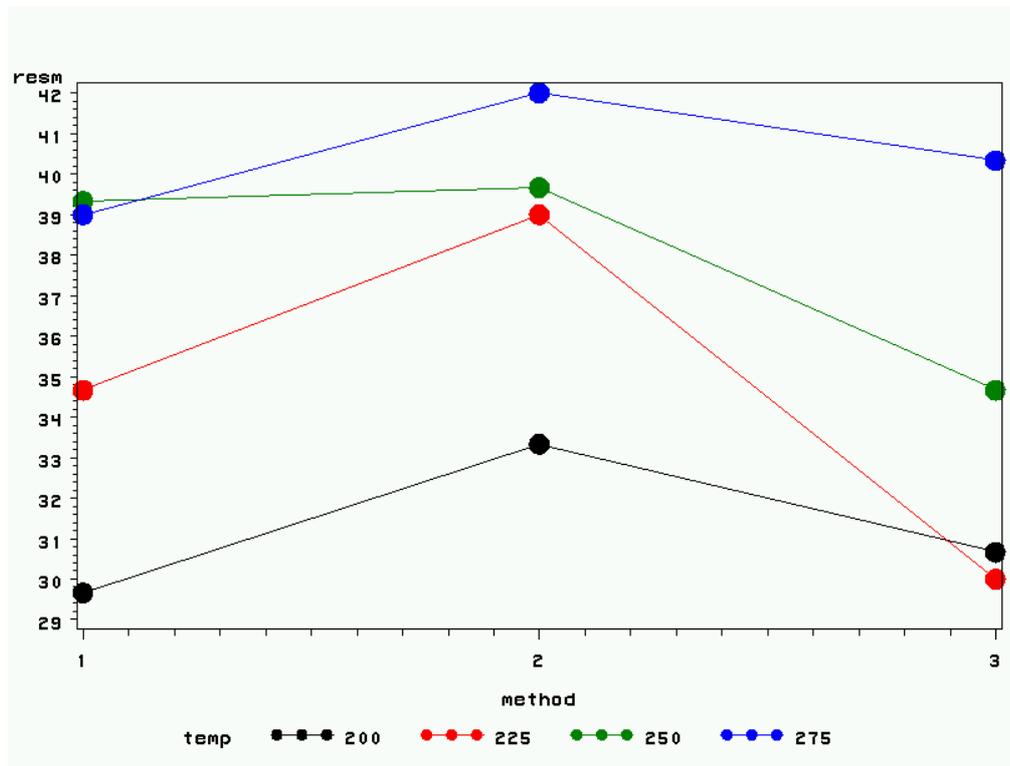


FIGURE 10.1: Graphique d'interaction pour l'expérience sur la pâte

```

34 41 38 42
29 26 33 36
28 32 40 41
31 36 42 40
31 30 32 40
31 37 41 40
35 40 39 44
32 34 39 45 ;

```

```

proc glm data=pate;
class bloc method temp;

```

```

model force=bloc|method temp temp*method;
test h=method e=bloc*method;
run;

```

La sortie de ce programme est:

Dependent Variable: force

Source	DF	Squares	Sum of Mean Square	F Value	Pr > F
Model	17	751.4722222	44.2042484	11.13	<.0001
Error	18	71.5000000	3.9722222		
Corrected Total	35	822.9722222			

R-Square	Coeff Var	Root MSE	force Mean
0.913120	5.531963	1.993043	36.02778

Source	DF	Type III SS	Mean Square	F Value	Pr > F
bloc	2	77.5555556	38.7777778	9.76	0.0013
method	2	128.3888889	64.1944444	16.16	<.0001
bloc*method	4	36.2777778	9.0694444	2.28	0.1003
temp	3	434.0833333	144.6944444	36.43	<.0001
method*temp	6	75.1666667	12.5277778	3.15	0.0271

Tests of Hypotheses Using the Type III MS for bloc*method as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
method	2	128.3888889	64.1944444	7.08	0.0485

On voit que TEMPÉRATURE est le facteur le plus important. On note que la variance intra de 3.97 est beaucoup plus faible que la variance inter (9.07). Lorsque testé avec le MS bloc*method, le facteur METHOD a un seuil observé de 4.8%. Il est significatif mais de justesse. La seule composante du modèle qui est fortement significative est TEMPÉRATURE; c'est bien ce que l'on voit sur le graphique d'interaction.

10.5 Expériences avec facteurs croisés et emboîtés

Pour présenter ce type de modèles, considérons une expérience en psychologie qui étudie l'impact de trois types de LEADERSHIP et de deux types de PROJETS sur Y une mesure de la performance d'une personne. L'expérience se déroule de la façon suivante. Disons que 15 personnes sont disponibles pour l'expérience. Au départ, on les divise au hasard en trois groupes de 5. Chaque groupe se voit attribuer un chef; un des groupes a un chef autoritaire, un autre un chef démocrate et le dernier un chef qui n'exerce pas son autorité. Le type de chef définit la modalité du facteur LEADERSHIP. Une fois les groupes formés, chaque membre se voit attribuer des projets qu'il doit mener à terme. Deux de ces PROJETS sont évalués, un simple et un complexe. Les données sont donc $30 = 15 \times 2$ évaluation de projets réalisés par les participants à l'expérience.

On reconnaît les deux facteurs fixes de l'expérience, LEADERSHIP et PROJET. Il y aussi un facteur aléatoire SUJET. On dispose de 15 sujets qui sont séparés au hasard en 3 groupes; chaque groupe est ensuite assigné à une modalité de LEADERSHIP. On dit que SUJET est *emboîté* dans LEADERSHIP car chaque modalité de LEADERSHIP reçoit son propre échantillon de SUJETS. Le facteur SUJET serait *croisé* avec LEADERSHIP si on obtenait, pour chaque

SUJET, des observations à chaque modalité de LEADERSHIP. Ce n'est pas le cas ici. Deux mesures sont prises sur chaque SUJET; on est donc en présence d'une *expérience à mesures répétées*.

Cette expérience est semblable à un schéma split-plot. Les SUJETS jouent le même rôle que les grandes parcelles. La table ANOVA va avoir deux parties: une composante inter-SUJET où on fait de l'inférence concernant le facteur LEADERSHIP et une partie intra-SUJET pour étudier PROJET et LEADERSHIP*PROJET. Il y a tout de même une différence importante avec un schéma split-plot: il n'y a pas de blocs dans une expérience à mesures répétées. De plus le facteur aléatoire est emboîté dans le facteur inter-SUJETS. On écrit ceci comme:

$$\text{SUJET}(\text{LEADERSHIP}).$$

Dans la construction du modèle on applique la règle suivante: si B (avec J modalités) est emboîté dans A (avec I modalités) on ne met pas d'interaction A*B dans le modèle. On aura une composante B(A) qui dit que B est emboîté dans A, avec $I(J-1)$ degrés de liberté.

Le modèle pour les données s'écrit:

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + \epsilon_{ijk},$$

avec

$i = 1, 2, 3, 5$ représente la modalité de LEADERSHIP

$j = 1, 2, 3, 5$ est associé aux SUJETS dans LEADERSHIP

$k = 1, 2$ représente la modalité de PROJET.

Les facteurs aléatoires sont , $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$ et $\epsilon_{ijk} \sim N(0, \sigma^2)$. Le paramètre σ_β^2 représente la variabilité inter-SUJETS. La table ANOVA pour ce modèle est donnée au tableau 10.2. On pourrait ajouter au modèle la composante $\gamma*\beta_{kj(i)}$ pour l'interaction PROJET*SUJET(LEADERSHIP). La somme de carrés pour ce terme coïnciderait alors avec la somme de carrés pour les erreurs dans le tableau 10.2.

Une règle utile pour calculer les degrés de liberté pour une somme de carrés avec termes croisés et emboîtés: il faut former un produit où une composante de la somme de carrés

contribue son nombre de modalités si un facteur est emboîté dans cette composante et son nombre de modalités moins 1 sinon.

Source	d.d.l.	Sommes de carrés	F
LEADER	$I - 1$	$KJ \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$	$MS_{LEADER} / MS_{SUJET(L)}$
SUJET(LEADER)	$I(J - 1)$	$K \sum_{i,j} (\bar{Y}_{ij.} - \bar{Y}_{i..})^2$	
PROJET	$K - 1$	$IJ \sum_i (\bar{Y}_{..K} - \bar{Y}_{...})^2$	MS_{PROJET} / MSE
LEADER*PROJET	$(K - 1)(I - 1)$	$J \sum_{i,k} (\bar{Y}_{i.k} - \bar{Y}_{..k} - \bar{Y}_{i..} + \bar{Y}_{...})^2$	$MS_{P * C} / MSE$
Erreur	$I(J - 1)(K - 1)$	$\sum_{i,j,k} (\bar{Y}_{ijk} - \bar{Y}_{i.k} - \bar{Y}_{ij.} + \bar{Y}_{i..})^2$	

TABLEAU 10.2: Table ANOVA pour un schéma avec mesures répétées (note LEADERSHIP est parfois écrit LEADER ou L)

À faire: Calculer les espérances des sommes de carrés moyens pour le modèle avec mesures répétées.

Le programme SAS pour analyser les données est le suivant:

```
data psych; do leader=1 to 3; do sujet=1 to 5; do projet=1 to 2;
input eval @; output; end; end; end;
datalines;
5 0
6 2
7 1
5 0
7 2
7 7
5 6
7 7
6 7
7 7
```

```

4 2
3 3
5 4
3 2
3 3 ;
proc format; value fleader 1='autoritaire' 2='democrate'
3='laisser faire'; value fprojet 1='simple' 2='complexe';run;

proc glm data=psych;
format leader fleader. projet fprojet.;
class leader sujet projet;
model eval=leader sujet(leader) projet leader*projet;
test h=leader e=sujet(leader);
slice leader*projet/ by leader; run;

proc sort data=psych; by leader projet; run; proc means data=psych;
by leader projet; var eval; output out=moy mean=resm;proc print
data=moy;run; symbol1 interpol=join
value=dot
height=2;
proc gplot data=moy; format leader fleader. projet fprojet.; plot
resm*projet=leader;run;

```

La sortie SAS est:

Dependent Variable: eval

Source	DF	Squares	Sum of Mean Square	F Value	Pr > F
Model	17	150.3666667	8.8450980	35.38	<.0001

Error	12	3.0000000	0.2500000
-------	----	-----------	-----------

R-Square	Coeff Var	Root MSE	eval Mean
0.980439	11.27820	0.500000	4.433333

Source	DF	Type I SS	Mean Square	F Value	Pr> F
leader	2	70.86666667	35.43333333	141.73	<.0001
sujet(leader)	12	15.00000000	1.25000000	5.00	0.0046
projet	1	24.30000000	24.30000000	97.20	<.0001
leader*projet	2	40.20000000	20.10000000	80.40	<.0001

Tests of Hypotheses Using sujet(leader) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr> F
leader	2	70.86666667	35.43333333	28.35	<.0001

The GLM Procedure
Least Squares Means

leader	projet	eval LSMEAN
autoritaire	complexe	1.00000000

autoritaire	simple	6.00000000
democrate	complexe	6.80000000
democrate	simple	6.40000000
laisser faire	complexe	2.80000000
laisser faire	simple	3.60000000

leader*projet Effect Sliced by leader for eval

leader	DF	Squares	Mean Square	F Value	Pr > F
autoritaire	1	62.500000	62.500000	250.00	<.0001
democrate	1	0.400000	0.400000	1.60	0.229
laisser faire	1	1.600000	1.600000	6.40	0.0264

On note donc que les deux modalités de PROJET ne sont pas différentes pour un leadership démocratique, elles le sont marginalement pour une leadership laisser-faire et fortement pour un leadership autoritaire. Dans les deux derniers ce sont les PROJETS simples qui reçoivent la meilleure évaluation.

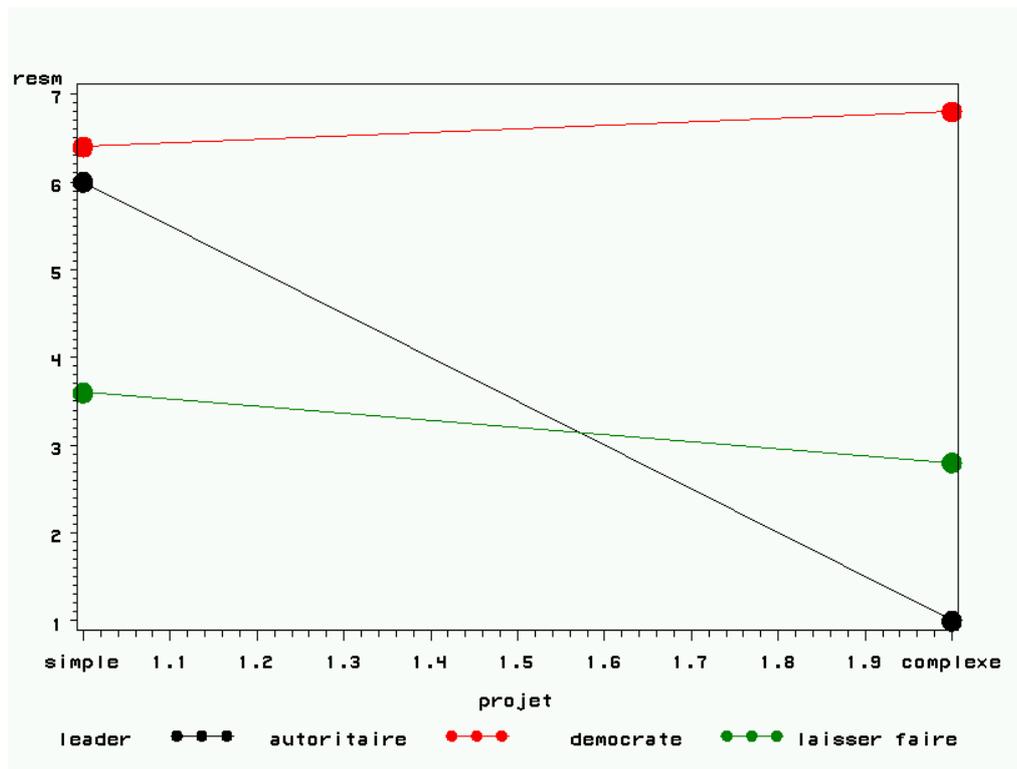


FIGURE 10.2: Graphique d'interaction