

CHAPITRE 8

ANOVA à deux facteurs fixes

Dans un grand nombre de situations expérimentales, plusieurs facteurs interviennent dans la définition des traitements. Par exemple:

- l'effet du produit et de sa concentration sur les résultats;
- l'effet du type du sol et de la température sur le rendement d'une plante dans une serre.

Dans ce chapitre, on s'intéresse à analyser des expériences visant à étudier les effets combinés de plusieurs facteurs sur une même variable. Nous nous limitons cependant aux facteurs fixes, c'est à dire où les modalités eux-mêmes sont d'intérêt. Dans ce contexte un traitement est une combinaison de modalités des facteurs à l'étude. Par exemple dans une expérience à deux facteurs, le traitement (1,2) est appliquée à une unité si elle reçoit la modalité 1 du premier facteur et la modalité 2 du deuxième.

8.1 Les cas des facteurs dichotomiques

Supposons que l'on s'intéresse à deux facteurs A et B , chacun avec deux modalités, 1 et 2. Il y a quatre traitements possibles, (1,1), (1,2), (2,1), et (2,2). On observe des échantillons de taille n pour chaque traitement. On peut traiter ces données avec un modèle d'ANOVA à un facteur. La somme de carrés inter traitement, SS_{Bet} , a trois degré de liberté.

Appelons $\bar{Y}_{11.}, \bar{Y}_{12.}, \bar{Y}_{21.}, \bar{Y}_{22.}$ les quatre moyennes à l'étude. On peut décomposer $SSBet$ à l'aide de contrastes orthogonaux. Le contraste qui compare les deux modalités du premier facteur est

$$c_A = \mu_{11} + \mu_{12} - \mu_{21} - \mu_{22}.$$

Celui qui compare les deux modalités du deuxième facteur s'écrit

$$c_B = \mu_{11} - \mu_{12} + \mu_{21} - \mu_{22}.$$

Notez que les contrastes pour c_A et c_B sont orthogonaux. La somme de carrés inter échantillon a trois degrés de liberté. On peut décomposer cette somme à l'aide de trois contrastes orthogonaux. On a déjà c_A et c_B , le troisième que l'on appellera c_{AB} s'écrit

$$c_{AB} = \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}.$$

On vérifie facilement que ce contraste est orthogonal aux deux premiers. En mots, c_{AB} est la différence entre une mesure, $\mu_{11} - \mu_{12}$, de l'effet du facteur B à la modalité 1 du facteur A et la mesure correspondante $\mu_{21} - \mu_{22}$ à la modalité 2 de A . Si c_{AB} est non nul, l'effet du facteur B varie selon la modalité du facteur A . De même l'effet de A dépend de la modalité de B . **On dit qu'il y a une interaction entre les deux facteurs.**

En fait les coefficients $(1,-1,-1,1)$ de c_{AB} sont égaux au produit terme à terme des coefficients de c_A , $(1,1,-1,-1)$, par ceux de c_B , $(1,-1,1,-1)$. On peut donc dire que c_{AB} représente l'interaction entre c_A et c_B . Ces trois contrastes permettent de décomposer la somme de carrés pour la variabilité inter échantillon en trois parties, une mesurant l'effet du facteur A , une autre celui du facteur B et la troisième l'interaction entre ces deux facteurs:

$$\begin{aligned} SSBet &= n \frac{(\bar{Y}_{11.} + \bar{Y}_{12.} - \bar{Y}_{21.} - \bar{Y}_{22.})^2}{4} + n \frac{(\bar{Y}_{11.} - \bar{Y}_{12.} + \bar{Y}_{21.} - \bar{Y}_{22.})^2}{4} + n \frac{(\bar{Y}_{11.} - \bar{Y}_{12.} - \bar{Y}_{21.} + \bar{Y}_{22.})^2}{4} \\ SSBet &= SSA + SSB + SSAB. \end{aligned} \tag{8.1}$$

Cette décomposition n'est pas spécifique au cas où les facteurs sont dichotomiques. Si les facteurs A et B ont respectivement I et J modalités, il y a en tout IJ échantillons et la somme de carrés pour la variabilité inter échantillon peut s'écrire comme une somme de trois termes, voir (8.1). De plus étant donné un contraste c_A qui compare des modalités de A entre

elles et un contraste c_B qui compare des modalités de B , on peut construire un contraste d'interaction en faisant le produit termes à termes des coefficients de chacun des contrastes. Il y a donc $I - 1$ degrés de liberté associés aux différences entre les modalités de A , $J - 1$ aux modalités de B et $(I - 1)(J - 1)$ à l'interaction entre les deux facteurs. Si deux facteurs interagissent, le classement des modalités de l'un peut dépendre de la modalité de l'autre. Il faut donc comparer les modalités de A entre elles à chaque modalité de B , ou vice-versa.

8.2 Le modèle dans le cas général

Dans le cas où on a deux facteurs, le modèle se présente comme suit:

$$Y_{ijk} = \mu_{ij} + e_{ijk}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, n_{ij}, \quad (8.2)$$

où Y_{ijk} est la variable réponse, μ_{ij} la moyenne de la cellule (i, j) et n_{ij} la taille de l'échantillon présent à la cellule (i, j) . Dans ce chapitre, on ne considérera que les plans balancés, c'est à dire où tous les n_{ij} sont égaux. Soit n leur valeur commune. La taille totale de tous les échantillons combinés est $N = nIJ$.

Comme dans le modèle à un facteur fixe, les erreurs e_{ijk} sont supposées indépendantes et identiquement distribuées selon $N(\mu, \sigma^2)$. Plusieurs paramétrisations sont possibles pour les μ_{ij} . Ici, on choisit la paramétrisation suivante:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \quad i = 1, \dots, I, \quad j = 1, \dots, J. \quad (8.3)$$

Les paramètres α_i , $i = 1, \dots, I$ et β_j , $j = 1, \dots, J$ représentent les effets des modalités des facteurs A et B respectivement. Ils satisfont aux contraintes $\sum_{i=1}^I \alpha_i = 0$ et $\sum_{j=1}^J \beta_j = 0$. Les paramètres γ_{ij} représentent l'interaction entre ces deux facteurs. Ils satisfont aux contraintes $\sum_{i=1}^I \gamma_{ij} = 0$ pour $j = 1, \dots, J$ et $\sum_{j=1}^J \gamma_{ij} = 0$ pour $i = 1, \dots, I$. Sous ces conditions, on a:

$$Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2) \quad (8.4)$$

En absence d'interaction, c'est à dire lorsque $\gamma_{ij} = 0$ pour tout (i, j) , on dit qu'on a un modèle additif.

EXEMPLE 8.1 Une étude vise à évaluer l'impact de deux facteurs sur la rentabilité d'un commerce en détail: le type d'agglomération (centre urbain, petite ville ou zone rurale) et taille du commerce (petite, moyenne et grande). Une absence d'interaction peut se traduire ici par dire que la différence entre la rentabilité d'un grand et d'un moyen commerce (d'une même zone) ne dépend pas du type d'agglomération.

En d'autres termes, l'absence d'interaction se traduit par: $\forall(i, j)$, on a $\mu_{ij} - \mu_{ij'} = \mu_{i'j} - \mu_{i'j'}$. Lorsque les deux facteurs n'interagissent pas, la poursuite de l'analyse est simplifiée. En effet on étudie séparément l'effet des facteurs A et B sur la variable dépendante à l'aide des techniques vues en analyse de variance à un facteur (comparaisons multiples, contrastes).

Si les deux facteurs interagissent, il n'est pas opportun d'étudier les facteurs A et B séparément. En fait si A et B interagissent certains auteurs suggèrent de ne pas considérer les test pour les effets individuels de A et B . Il faut être prudent ici; en présence d'interaction les tests sur A et B peuvent identifier le facteur qui expliquent le plus les variations de Y . Une interaction, même significative, est peut-être peu importante par rapport à des effets marginaux fortement significatifs.

8.3 Distributions des moyennes

Définissons les différentes moyennes $\bar{Y}_{ij.}$, $\bar{Y}_{i..}$, $\bar{Y}_{.j.}$ et $\bar{Y}_{...}$ respectivement par:

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

$$= \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \bar{Y}_{ij}.$$

Pour un couple (i, j) fixé, les n variables aléatoires $Y_{ij1}, Y_{ij2}, \dots, Y_{ijn}$ sont indépendantes et identiquement distribuées selon $N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \sigma^2)$. On en déduit que la moyenne de la cellule \bar{Y}_{ij} est distribuée selon:

$$\bar{Y}_{ij} \sim N(\mu + \alpha_i + \beta_j + \gamma_{ij}, \frac{\sigma^2}{n})$$

Pour $i = 1, \dots, I$, $\bar{Y}_{i..}$ est une combinaison linéaire de variables aléatoires indépendantes de même variance mais d'espérances différentes. La moyenne de la ligne $\bar{Y}_{i..}$ est alors normalement distribuée. Son espérance et sa variance sont égales à:

$$\begin{aligned} E[\bar{Y}_{i..}] &= E\left[\frac{1}{J} \sum_{j=1}^J \bar{Y}_{ij}\right] \\ &= \frac{1}{J} \sum_{j=1}^J E[\bar{Y}_{ij}] \\ &= \frac{1}{J} \sum_{j=1}^J \{\mu + \alpha_i + \beta_j + \gamma_{ij}\} \\ &= \frac{1}{J} \left\{ \sum_{j=1}^J (\mu + \alpha_i) + \sum_{j=1}^J \beta_j + \sum_{j=1}^J \gamma_{ij} \right\} \\ &= \frac{1}{J} J(\mu + \alpha_i) = \mu + \alpha_i \quad \text{car} \quad \sum_{j=1}^J \beta_j = \sum_{j=1}^J \gamma_{ij} = 0 \quad \forall j = 1, \dots, J \end{aligned}$$

et

$$\begin{aligned} \text{Var}[\bar{Y}_{i..}] &= \text{Var}\left[\frac{1}{J} \sum_{j=1}^J \bar{Y}_{ij}\right] \\ &= \frac{1}{J^2} \sum_{j=1}^J \text{Var}[\bar{Y}_{ij}] \\ &= \frac{1}{J^2} \sum_{j=1}^J \frac{\sigma^2}{n} \end{aligned}$$

$$= \frac{1}{J^2} \frac{J\sigma^2}{n} = \frac{\sigma^2}{Jn}$$

respectivement.

On a alors

$$\bar{Y}_{i..} \sim N\left(\mu + \alpha_i, \frac{\sigma^2}{Jn}\right) \quad \forall i = 1, \dots, I.$$

De même, on montre que

$$\bar{Y}_{.j.} \sim N\left(\mu + \beta_j, \frac{\sigma^2}{In}\right) \quad \forall j = 1, \dots, J.$$

En suivant un raisonnement similaire pour la moyenne globale $\bar{Y}_{...}$, on obtient

$$\bar{Y}_{...} \sim N\left(\mu, \frac{\sigma^2}{N}\right).$$

8.4 Décomposition des sommes des carrés

La somme des carrés des erreurs SST ou *sum of squares total*, est définie par

$$SST = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^n \{Y_{ijk} - \bar{Y}_{...}\}^2.$$

En écrivant chaque terme $\{Y_{ijk} - \bar{Y}_{...}\}^2$ comme $\{(Y_{ijk} - \bar{Y}_{ij.}) + (\bar{Y}_{ij.} - \bar{Y}_{...})\}^2$ et en développant le carré, on obtient $SST = SSE + SS_{model}$ où

$$SSE = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^n \{Y_{ijk} - \bar{Y}_{ij.}\}^2$$

et

$$\begin{aligned} SS_{model} &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^n \{\bar{Y}_{ij.} - \bar{Y}_{...}\}^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J n \{\bar{Y}_{ij.} - \bar{Y}_{...}\}^2 \end{aligned}$$

Les sommes des produits $(Y_{ijk} - \bar{Y}_{ij.})(\bar{Y}_{ij.} - \bar{Y}_{...})$ sont nuls.

D'autre part, on peut décomposer SS_{model} selon

$$\begin{aligned}
 SS_{model} &= n \sum_{i=1}^I \sum_{j=1}^J \{\bar{Y}_{ij.} - \bar{Y}_{...}\}^2 \\
 &= n \sum_{i=1}^I \sum_{j=1}^J \{(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...})\}^2 \\
 &= n \sum_{i=1}^I \sum_{j=1}^J \{\bar{Y}_{i..} - \bar{Y}_{...}\}^2 + n \sum_{i=1}^I \sum_{j=1}^J \{\bar{Y}_{.j.} - \bar{Y}_{...}\}^2 + n \sum_{i=1}^I \sum_{j=1}^J \{\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\}^2
 \end{aligned}$$

Les différentes double sommes étant nulles, on a alors:

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

où

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

On obtient finalement $SST = SSA + SSB + SS(AB) + SSE$.

On définit ensuite les moyennes des carrées par

$$\begin{aligned}
 MSA &= \frac{SSA}{I-1} \\
 MSB &= \frac{SSB}{J-1} \\
 MS(AB) &= \frac{SS(AB)}{(I-1)(J-1)} \\
 MSE &= \frac{SSE}{IJ(n-1)}
 \end{aligned}$$

En utilisant des techniques de formes quadratiques, on montre le résultat suivant:

PROPOSITION 8.1 *Les statistiques MSA , MSB , $MS(AB)$ et MSE sont indépendantes et distribuées selon*

$$\begin{aligned}\frac{(I-1)MSA}{\sigma^2} &\sim \chi_{I-1}^2\left(\frac{Jn}{\sigma^2} \sum_{i=1}^I \alpha_i^2\right) \\ \frac{(J-1)MSB}{\sigma^2} &\sim \chi_{J-1}^2\left(\frac{In}{\sigma^2} \sum_{j=1}^J \beta_j^2\right) \\ \frac{(I-1)(J-1)SS(AB)}{\sigma^2} &\sim \chi_{(I-1)(J-1)}^2\left(\frac{n}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^J \gamma_{ij}^2\right) \\ \frac{IJ(n-1)MSE}{\sigma^2} &\sim \chi_{IJ(n-1)}^2\end{aligned}$$

On en déduit que

$$\frac{IJ(n-1)MS_{model}}{\sigma^2} \sim \chi_{IJ-1}^2\left(\frac{n}{\sigma^2} \sum_{i,j} (\mu_{ij} - \mu)^2\right)$$

où $MS_{model} = SS_{model}/(IJ-1)$ et $\mu = (\sum_{i,j} \mu_{ij})/IJ$.

Dans le cadre d'une ANOVA à deux facteurs, plusieurs tests sont possibles.

En absence d'interaction, les termes $\{\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\}^2$ ont tendance à être petits. On rejette l'absence d'une interaction entre les facteurs A et B si $MS(AB)$ est élevé, ou si $F_{AB} = MS(AB)/MSE$ est élevé. Or sous H_0 , on a $\gamma_{ij} = 0$ pour tout couple (i, j) . $MS(AB)$ possède alors une loi de khi-deux centrée et $F_{AB} \sim F_{(I-1)(J-1), IJ(n-1)}$. On rejette alors H_0 si $F_{AB} > F_{(I-1)(J-1), IJ(n-1), 1-\alpha}$.

Si le facteur A n'a pas d'effets, les termes $\{\bar{Y}_{i..} - \bar{Y}_{...}\}^2$ ont tendance à être petits. On rejette l'absence d'un effet du facteur A si MSA est élevé, ou si $F_A = MSA/MSE$ est élevé. Or sous H_0 , on a $\alpha_i = 0$ pour tout i . MSA possède alors une loi de khi-deux centrée et $F_A \sim F_{I-1, IJ(n-1)}$. On rejette alors H_0 si $F_A > F_{I-1, IJ(n-1), 1-\alpha}$.

De même, on rejette l'absence d'un effet du facteur B si $F_B > F_{J-1, IJ(n-1), 1-\alpha}$ où $F_B = MSB/MSE \sim F_{J-1, IJ(n-1)}$ sous H_0 .

8.5 Estimation des paramètres

A partir de la propriété $(IJ(n-1)MSE)/\sigma^2 \sim \chi_{IJ(n-1)}^2$, on déduit un estimateur ponctuel $\hat{\sigma}^2 = MSE$ et par intervalle pour σ^2 :

$$\left[\frac{IJ(n-1)MSE}{\chi_{IJ(n-1),\alpha/2}^2}, \frac{IJ(n-1)MSE}{\chi_{IJ(n-1),1-\alpha/2}^2} \right]$$

Pour le paramètre μ , on sait que $\bar{Y}_{...} \sim N(\mu, \sigma^2/N)$ qu'on aussi écrire

$$\frac{\bar{Y}_{...} - \mu}{\sqrt{\frac{\sigma^2}{N}}} \sim N(0, 1).$$

On en déduit que:

$$\frac{\bar{Y}_{...} - \mu}{\sqrt{\frac{MSE}{N}}} \sim t_{IJ(n-1)}$$

L'estimation ponctuelle et par intervalle de confiance pour μ s'en suit: $\hat{\mu} = \bar{Y}_{...}$ et un I.C pour μ à $1 - \alpha$ s'écrit:

$$\left[\bar{Y}_{...} - t_{IJ(n-1),\alpha/2} \sqrt{\frac{MSE}{N}}, \bar{Y}_{...} + t_{IJ(n-1),\alpha/2} \sqrt{\frac{MSE}{N}} \right]$$

Pour les paramètres α_i , à partir de

$$\begin{aligned} \bar{Y}_{i..} &\sim N\left(\mu + \alpha_i, \frac{\sigma^2}{Jn}\right) \\ \bar{Y}_{...} &\sim N\left(\mu, \frac{\sigma^2}{N}\right), \end{aligned}$$

on peut déduire que $\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}$ peut être u estimateur sans biais de α_i . Au fait c'est l'estimateur de maximum de vraisemblance de α_i . Des calculs similaires à ceux utiliser à la page 36 nous donnent:

$$\bar{Y}_{i..} - \bar{Y}_{...} \sim N\left(\alpha_i, \frac{\sigma^2}{Jn} \left(\frac{I-1}{I}\right)\right)$$

On en déduit que:

$$\frac{\bar{Y}_{i..} - \bar{Y}_{...} - \alpha_i}{\sqrt{\frac{\sigma^2}{Jn} \left(\frac{I-1}{I}\right)}} \sim N(0, 1)$$

et que

$$\frac{\bar{Y}_{i..} - \bar{Y}_{...} - \alpha_i}{\sqrt{\frac{MSE}{Jn} \left(\frac{I-1}{I}\right)}} \sim t_{IJ(n-1)}.$$

L'I.C. pour α_i au niveau dwe confiance $1 - \alpha$ s'écrit alors:

$$[\bar{Y}_{i..} - \bar{Y}_{...} - t_{IJ(n-1), 1-\alpha/2} \sqrt{\frac{MSE}{Jn} \frac{I-1}{I}}, \bar{Y}_{i..} - \bar{Y}_{...} + t_{IJ(n-1), 1-\alpha/2} \sqrt{\frac{MSE}{Jn} \frac{I-1}{I}}]$$

De même, on montre qu'un I.C. pour β_j peut s'écrire:

$$[\bar{Y}_{.j.} - \bar{Y}_{...} - t_{IJ(n-1), 1-\alpha/2} \sqrt{\frac{MSE}{In} \frac{J-1}{J}}, \bar{Y}_{.j.} - \bar{Y}_{...} + t_{IJ(n-1), 1-\alpha/2} \sqrt{\frac{MSE}{In} \frac{J-1}{J}}]$$

Finalement, pour γ_{ij} , on montre, en utilisant les mêmes techniques, que

$$\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...} \sim N\left(\gamma_{ij}, \frac{\sigma^2 (I-1)(J-1)}{nIJ}\right)$$

On en déduit un I.C. pour γ_{ij} au niveau $a - \alpha$

$$[\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...} \pm t_{IJ(n-1), \alpha/2} \sqrt{\frac{MSE}{n} \frac{(I-1)(J-1)}{IJ}}]$$

8.6 Validation des hypothèses

La validation des hypothèses se fait comme pour l'ANOVA à un facteur car l'ANOVA à deux facterus est un cas particulier de l'ANOVA à un facteur. La validation d'hypothèses s'effectue sur les résidus. Ces derniers sont définis par

$$\hat{e}_{ijk} = Y_{ijk} - (\hat{\mu} + \hat{\alpha}_j + \hat{\beta}_j + \hat{\gamma}_{ij}).$$

8.7 Exemple

On a planifié une expérience pour étudier le voltage maximum d'une pile en fonction du matériel utilisé pour faire la pile (facteur *MAT* avec $I = 3$ modalités) et de la température ambiante (facteur *TEMP* avec $J = 3$ modalités). Des échantillons de taille $n = 4$ sont observés pour chaque traitement.

Programme SAS de lecture des données.

```

data voltage;
do mat= 1 to 3;
do temp= 1 to 3;
do rep = 1 to 4;
input volt @; output; end;end;end;
datalines;
130 155 74 180
34 40 80 75
20 70 82 58
150 188 159 126
136 122 106 115
25 70 58 45
138 110 168 160
174 120 150 139
96 104 82 60
;
    data voltage2; set voltage;
tempn=50; if temp=2 then tempn=65; if temp=3 then tempn=80; drop
temp rep; run;

```

Pour évaluer visuellement si les facteurs interagissent on fait le **graphique d'interaction**. On fixe les modalités d'un facteur, disons MAT . Pour chaque modalité i de MAT , on fait la ligne brisée de (j, \bar{Y}_{ij}) . On a alors trois lignes brisées. Si ces lignes sont à peu près parallèles, il n'y a pas d'interaction. En effet dans ce cas $\bar{Y}_{ij} - \bar{Y}_{kj}$, la différence entre les modalités i et k de MAT , est à peu près la même pour toutes les modalités j de $TEMP$.

Le code SAS pour faire ce graphique est donné par

```

proc sort data=voltage2;
by mat tempn; run;
proc means data=voltage2;
by mat tempn;

```

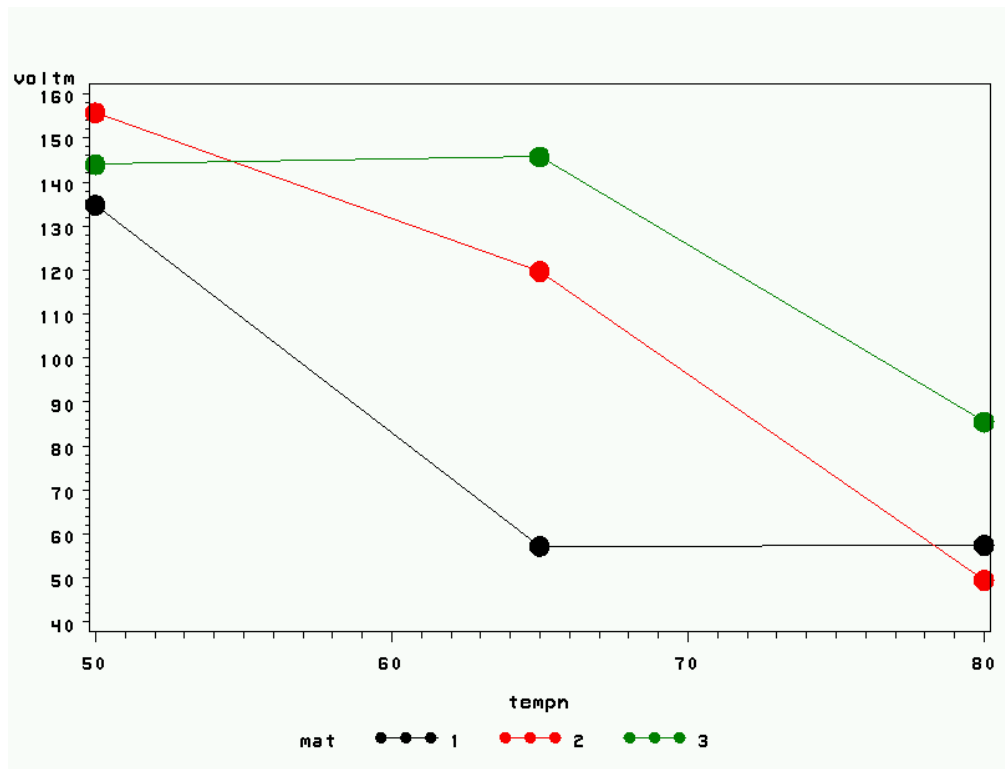


FIGURE 8.1: Graphique d'interaction

```
output out=moy mean=voltm;proc print data=moy;run;
```

```
symbol1 interpol=join
value=dot
height=2;
```

```
proc gplot data=moy; plot voltm*tempn=mat;run;
```

On remarque que les lignes sont à peu près parallèle sauf à la température 65. Cette interaction est-elle significative. Pour poursuivre l'analyse on peut calculer la table ANOVA, avec GLM. Le programme SAS est

```
proc glm data=voltage2;
```

```

class tempn mat;
model volt= tempn mat tempn*mat;
lsmeans tempn*mat / slice=tempn; run;

```

Dependent Variable: volt

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	59416.22222	7427.02778	11.00	<.0001
Error	27	18230.75000	675.21296		
Corrected Total	35	77646.97222			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
tempn	2	39118.72222	19559.36111	28.97	<.0001
mat	2	10683.72222	5341.86111	7.91	0.0020
tempn*mat	4	9613.77778	2403.44444	3.56	0.0186

L'énoncé slice permet de faire des tables ANOVA conditionnelles qui comparent les modalités de *MAT* à chaque température (on conditionn sur le facteur qui explique le plus de variabilité). Le résultat confirme ce que l'on voyait à la figure 1. C'est seulement à 65 degrés que les modalités de *MAT* sont différentes!

tempn*mat Effect Sliced by tempn for volt

tempn	DF	Sum of Squares	Mean Square	F Value	Pr > F
50	2	886.166667	443.083333	0.66	0.5269

65	2	16553	8276.333333	12.26	0.0002
80	2	2858.666667	1429.333333	2.12	0.1400

8.8 Modèle sans interaction

En absence d'interaction, le modèle s'écrit

$$Y_{ijk} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$$

Sous ce modèle, on a

$$SST = SSA + SSB + SSE'$$

où $SSE' = SSE + SS(AB)$. On dit que le terme d'interaction est absorbé par l'erreur. La nouvelle erreur est alors distribuée selon

$$\frac{(nIJ - I - J + 1)MSE'}{\sigma^2} \sim \chi_{nIJ - I - J + 1}^2$$

où $MSE' = SSE' / (nIJ - I - J + 1)$.

Les tests des effets des facteurs A et B se font alors par l'intermédiaire des statistiques F'_A et F'_B définies par $F'_A = MSA/MSE'$ et $F'_B = MSB/MSE'$ respectivement. Sous H_0 : pas d'effet du facteur A équivalent à $H_0 = \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$, $F'_A \sim F_{I-1, nIJ - I - J + 1}$ et on rejette H_0 si $F'_A > F_{I-1, nIJ - I - J + 1, 1-\alpha}$.

De même, on rejette l'existence d'un effet du facteur B si $F'_B > F_{J-1, nIJ - I - J + 1, 1-\alpha}$.

8.9 Modèle avec une observation par cellule

Dans plusieurs situations, des contraintes (coût, temps, ...etc) obligent les expérimentateurs à n'avoir qu'une observation par cellule ($n = 1$). Dans ce cas, on a $SSE = 0$ et $SST = SSA + SSB + SS(AB)$. C'est un modèle saturé sans erreur donc on ne peut pas effectuer de tests, ni trouver des intervalles de confiance.

Néanmoins, si on a des connaissances antérieures que l'interaction entre les facteurs n'est pas significative, on utilise alors le terme $SS(AB)$ comme terme d'erreur à la place de SSE .

Ainsi, on estime la variance σ^2 par $MS(AB)$ et on teste l'effet des facteurs A et B en utilisant les statistiques $F''_A = MSA/MS(AB)$ et $F''_B = MSB/MS(AB)$ respectivement. En absence d'effet du facteur A , $F''_A \sim F_{I-1, (I-1)(J-1)}$. Ainsi, on rejette H_0 si $F''_A > F_{I-1, (I-1)(J-1), 1-\alpha}$. De même, on rejette l'absence d'un effet du facteur B lorsque $F''_B > F_{J-1, (I-1)(J-1), 1-\alpha}$.