Analyse de variance avec interaction

F. Husson husson@agrocampus-ouest.fr

Données - exemple

Exemple: 2 produits; 3 juges; 2 répétitions

	Juge 1	Juge 2	Juge 3	Moy
Produit 1	1	1	2 4	2
Produit 2	2	4 6	4 6	4
Moy	2	3	4	3

Prod.	Juge	Note		
P1	J1	1		
P1	J1	3		
P1	J2	1		
P1	J2	1		
P1	J3	2		
P1	J3	4		
P2	J3	6		

Données - notations

					_			
Séance	Juge	Produit	Sucre	Acide	Amer	Cacao	Lait	
S1	J1	P6	4	3	2	5.5	7.5	
S1	J1	P4	1.2	4.4	6	7.6	5.5	
S1	J1	P2	1.8	3	2.6	5	2.4	
S1	J1	P5	1.5	3.5	7.1	7.5	7.3	
S1	J1	P1	1	5.5	9.3	8.6	8.1	
S1	J1	P3	9	1	0	0.5	3.7	
:	:		÷	:	:	:	:	
S1	J2	P5	3.9	2	2.4	5.6	4.8	
S1	J2	P6	2.4	4	4.9	5.3	5.8	

Questions

Y a-t-il des différences d'amertume entre chocolats ?

Les juges utilisent-ils l'échelle de note de la même façon ?

L'amertume des chocolats estelle évaluée de la même façon d'une séance à l'autre ?

Les juges évaluent-ils les chocolats de la même façon ?

2

Données - notations

- · Y variable quantitative
- F1, F2, ... variables qualitatives à I, J, ... modalités
- n_{ii} répétitions pour le couple (i,j)

obs	F1	F2	\overline{y}		
1	1	1	<i>y</i> ₁₁₁		
i	:	:	:		
1	1	$y_{11n_{11}}$			
÷	÷	:	:		
i	j	y_{ijk}			
i	:	÷	÷		
n	I	J	$y_{IJn_{IJ}}$		

$$y_{ij\bullet} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$$

$$y_{i\bullet\bullet} = \frac{1}{n_i} \sum_{j,k} y_{ijk}$$

$$y_{\bullet j\bullet} = \frac{1}{n_j} \sum_{i,k} y_{ijk}$$

$$y_{\bullet \bullet \bullet} = \frac{1}{n_i} \sum_{i,j,k} y_{ijk}$$

4

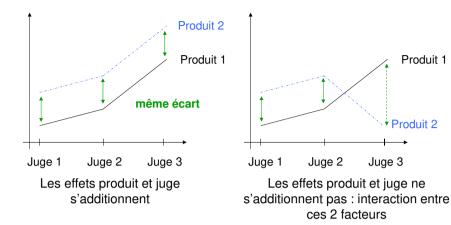
Questions

- Y a-t-il un effet « produit » sur la note ?
- Y a-t-il un effet « juge » sur la note ?
- Y a-t-il une interaction entre les deux facteurs ?

Tests

• Décision dans l'incertain : notion de test

Définition de l'interaction



Interaction:

5

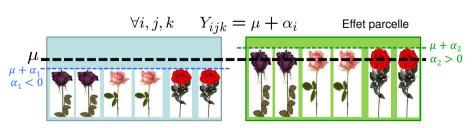
l'effet d'un facteur sur Y diffère selon les modalités de l'autre

6

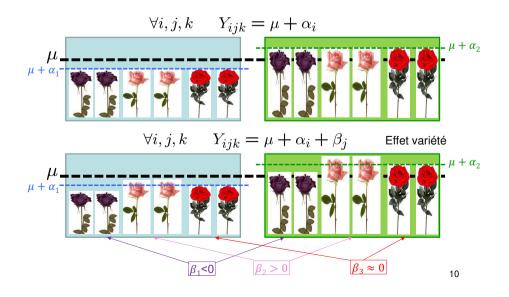
Exemple : 2 parcelles, 3 variétés de rose

$$i=1,...,2$$
 $j=1,...,3$ $k=1,2$

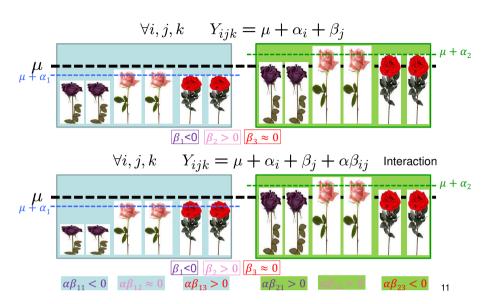
$$\mu = \mu$$



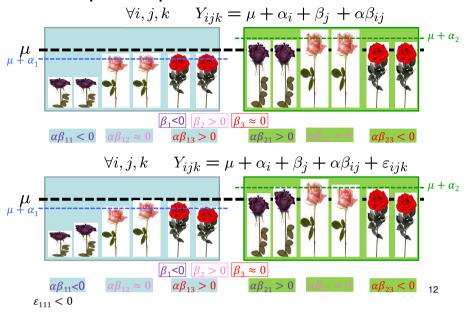
Exemple : 2 parcelles, 3 variétés de rose



Exemple: 2 parcelles, 3 variétés de rose

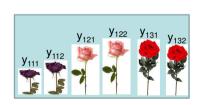


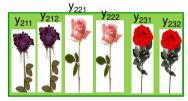
Exemple : 2 parcelles, 3 variétés de rose



Exemple : 2 parcelles, 3 variétés de rose

On ne dispose que de l'info suivante : la taille de chaque fleur, sa variété et la parcelle dans laquelle elle est cultivée





Et on cherche à estimer :

 μ : la taille moyenne des fleurs (quelle que soit la parcelle et la variété)

 α_i : l'effet de la parcelle i

 β_i : l'effet de la variété j

 $\alpha \beta_{ij}$: l'effet de l'interaction variété - parcelle

Définition du modèle à 2 facteurs

 μ effet moyen

 $lpha_i$ effet principal du niveau i du facteur 1

 eta_{j} effet principal du niveau j du facteur 2

 $lphaeta_{ij}$ effet de l'interaction des facteurs 1 et 2 pour les niveaux i et j

 $arepsilon_{ijk}$ résiduelle

Écriture matricielle du modèle :

$$Y = X\beta + E$$
 avec $\mathbb{E}(E) = 0$ et $\mathbb{V}(E) = \sigma^2 Id$

Contraintes

1 + I + J + IJ paramètres mais IJ paramètres indépendants modèle sur-paramétré besoin de contraintes

Contraintes:

$$\begin{array}{ll} \alpha_1=0 & \text{La modalité 1 du 1}^{\text{er}} \text{ facteur sert de référence} \\ \beta_1=0 & \text{La modalité 1 du 2}^{\text{ème}} \text{ facteur sert de référence} \\ \hline \begin{cases} \forall \ i,\alpha\beta_{i1}=0 \\ \forall \ j,\alpha\beta_{ij}=0 \end{cases} & \text{Les interactions avec les modalités 1 du 1}^{\text{er}} \text{ et du 2}^{\text{ème}} \text{ facteur servent de référence} \\ \end{cases}$$



Il est extrêmement difficile d'interpréter les coefficients estimés avec ces contraintes pour des modèles avec interactions (l'interprétation d'un coefficient dépend du modèle)

ATTENTION: ces contraintes sont utilisées par défaut dans R

→ ne pas utiliser les fonctions par défaut

16

Contraintes

1 + I + J + IJ paramètres mais IJ paramètres indépendants modèle sur-paramétré besoin de contraintes

Contraintes:

$$\sum_{i=1}^{I} \alpha_i = 0$$

$$\sum_{j=1}^{J} \beta_j = 0$$

$$\forall i, \sum_{j=1}^{J} \alpha \beta_{ij} = 0$$

$$\forall j, \sum_{i=1}^{I} \alpha \beta_{ij} = 0$$

Exemple:

$$\begin{array}{c} Y_{111} = \mu \ + \ \alpha_1 \ + \beta_1 \ + \alpha\beta_{11} \ + \ \epsilon_{111} \\ Y_{112} = \mu \ + \ \alpha_1 \ + \beta_1 \ + \alpha\beta_{11} \ + \ \epsilon_{112} \\ Y_{121} = \mu \ + \ \alpha_1 \ + \beta_2 \ + \alpha\beta_{12} \ + \ \epsilon_{121} \\ Y_{122} = \mu \ + \ \alpha_1 \ + \beta_2 \ + \alpha\beta_{12} \ + \ \epsilon_{121} \\ Y_{131} = \mu \ + \ \alpha_1 \ - \beta_1 \ - \ \beta_2 \ - \alpha\beta_{11} \ - \ \alpha\beta_{12} \ + \ \epsilon_{131} \\ Y_{132} = \mu \ + \ \alpha_1 \ - \beta_1 \ - \ \beta_2 \ - \alpha\beta_{11} \ - \ \alpha\beta_{12} \ + \ \epsilon_{132} \\ Y_{211} = \mu \ - \ \alpha_1 \ + \beta_1 \ - \alpha\beta_{11} \ + \ \epsilon_{211} \\ Y_{212} = \mu \ - \ \alpha_1 \ + \beta_1 \ - \alpha\beta_{11} \ + \ \epsilon_{212} \\ Y_{221} = \mu \ - \ \alpha_1 \ + \beta_1 \ - \alpha\beta_{11} \ + \ \epsilon_{212} \\ Y_{221} = \mu \ - \ \alpha_1 \ + \beta_2 \ - \alpha\beta_{12} \ + \ \epsilon_{221} \\ Y_{222} = \mu \ - \ \alpha_1 \ + \beta_2 \ - \alpha\beta_{12} \ + \ \epsilon_{222} \\ Y_{231} = \mu \ - \ \alpha_1 \ - \beta_1 \ - \beta_2 \ + \alpha\beta_{11} \ + \ \alpha\beta_{12} \ + \ \epsilon_{232} \\ Y_{232} = \mu \ - \ \alpha_1 \ - \beta_1 \ - \beta_2 \ + \alpha\beta_{11} \ + \ \alpha\beta_{12} \ + \ \epsilon_{232} \\ Y_{231} = \mu \ - \ \alpha_1 \ - \beta_1 \ - \beta_2 \ + \alpha\beta_{11} \ + \ \alpha\beta_{12} \ + \ \epsilon_{232} \\ Y_{231} = \mu \ - \ \alpha_1 \ - \beta_1 \ - \beta_2 \ + \alpha\beta_{11} \ + \ \alpha\beta_{12} \ + \ \epsilon_{232} \\ Y_{231} = \mu \ - \ \alpha_1 \ - \beta_1 \ - \beta_2 \ + \alpha\beta_{11} \ + \ \alpha\beta_{12} \ + \ \epsilon_{232} \\ Y_{231} = \mu \ - \ \alpha_1 \ - \beta_1 \ - \beta_2 \ + \alpha\beta_{11} \ + \ \alpha\beta_{12} \ + \ \alpha\beta_{12} \ + \ \alpha\beta_{13} \ + \ \alpha\beta_{12} \ + \ \alpha\beta_{13} \ + \ \alpha\beta_{13} \ + \ \alpha\beta_{13} \ + \ \alpha\beta_{13} \ + \ \alpha\beta_{12} \ + \ \alpha\beta_{13} \ + \ \alpha\beta_{13} \ + \ \alpha\beta_{12} \ + \ \alpha\beta_{13} \ + \ \alpha\beta_{12} \ + \ \alpha\beta_{13} \ + \ \alpha$$

Estimation des paramètres du modèle

par la méthode des moindres carrés

Minimiser
$$\sum_{i,j,k} \varepsilon_{ijk}^2$$
 revient à minimiser $E^2 = (Y - X\beta)'(Y - X\beta)$
 $E^2 = Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$

Annulons la dérivée par rapport à β pour trouver le minimum

$$\frac{\partial (E^2)}{\partial \beta} = \frac{\partial (Y'Y)}{\partial \beta} - \frac{\partial (Y'X\beta)}{\partial \beta} - \frac{\partial (\beta'X'Y)}{\partial \beta} + \frac{\partial (\beta'X'X\beta)}{\partial \beta} = 0$$

$$0 - (X'Y) - (X'Y) + X'X\beta + X'X\beta = 0$$

$$X'Y = X'X\beta$$
Règles de calcul pour dérivation matricielle
$$\frac{\partial (X'A)}{\partial A} = \frac{\partial (A'X)}{\partial A} = X$$

Règles de calcul pour dérivation matricielle
$$\frac{\partial (X'A)}{\partial A} = \frac{\partial (A'X)}{\partial A} = X$$
$$\frac{\partial (A'X')}{\partial A} = \frac{\partial (AX)}{\partial A} = X'$$

X'X non inversible : sur-paramétrisation (besoin de contraintes) X'X inversible : l'estimateur de β est : $\hat{\beta} = (X'X)^{-1}X'Y$

Propriétés :
$$\mathbb{E}(\widehat{\beta})=\beta$$
 $\mathbb{V}(\widehat{\beta})=\sigma^2\;(X'X)^{-1}$
$$\widehat{\beta}-\beta\sim N(0,\sigma^2\;(X'X)^{-1})$$
 19

Estimation des paramètres du modèle

Prédiction et résidus

Valeurs prédites : $\hat{Y} = X\hat{\beta}$

$$\widehat{y}_{ijk} = \widehat{\mu} + \widehat{\alpha}_i + \widehat{\beta}_j + \widehat{\alpha\beta}_{ij} = y_{ij\bullet}$$

Résidus : $E = Y - \hat{Y}$

 $e_{ijk} = y_{ijk} - \hat{y}_{ijk}$

Estimateur de la variabilité résiduelle

$$\hat{\sigma}^2 = \frac{\sum_{i,j,k} e_{ijk}^2}{\dots}$$

Propriété : $\mathbb{E}(\hat{\sigma}^2) = \sigma^2$

20

Estimation des paramètres du modèle

Cas particulier du plan équilibré (complet équirépété) :

$$\widehat{\alpha}_i = y_{\bullet \bullet \bullet}$$

$$\widehat{\alpha}_i = y_{i \bullet \bullet} - y_{\bullet \bullet \bullet}$$

$$\widehat{\beta}_j = y_{\bullet j \bullet} - y_{\bullet \bullet \bullet}$$

$$\widehat{\alpha \beta}_{ij} = y_{ij \bullet} - y_{i \bullet \bullet} - y_{\bullet j \bullet} + y_{\bullet \bullet \bullet}$$

Décomposition de la variabilité

Cas complet et équirépété : les SC s'additionnent

$$\sum_{i,j,k} (Y_{ijk} - Y_{\bullet \bullet \bullet})^2 = \sum_{i,j,k} (Y_{i\bullet \bullet} - Y_{\bullet \bullet \bullet})^2 \qquad I-1$$

$$+ \sum_{i,j,k} (Y_{\bullet j \bullet} - Y_{\bullet \bullet \bullet})^2 \qquad J-1$$

$$+ \sum_{i,j,k} (Y_{ij\bullet} - Y_{i\bullet \bullet} - Y_{\bullet j \bullet} + Y_{\bullet \bullet \bullet})^2 \qquad (I-1)(J-1)$$

$$+ \sum_{i,j,k} (Y_{ijk} - Y_{ij\bullet})^2 \qquad ddl_R$$

$$SC_T = SC_A + SC_B + SC_{AB} + SC_B$$

Test global d'un effet

Test du facteur A:

Hypothèses :
$$H_0$$
 : $\forall i, \ \alpha_i = 0$ H_1 : $\exists i \ / \ \alpha_i \neq 0$

$$H_1: \exists i / \alpha_i \neq 0$$

$$\mathbb{E}(CM_A) = \sigma^2 + \frac{KJ}{I-1} \sum_i \alpha_i^2$$

$$\mathbb{E}(CM_B) = \sigma^2 + \frac{KI}{J-1} \sum_j \beta_j^2$$

$$\mathbb{E}(CM_{AB}) = \sigma^2 + \frac{K}{(I-1)(J-1)} \sum_{ij} \alpha \beta_{ij}^2$$

$$\mathbb{E}(CM_R) = \sigma^2$$

Idée:....

23

Test global d'un effet

Test du facteur A:

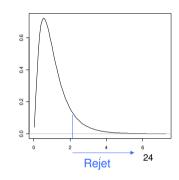
Hypothèses: H_0 : $\forall i, \alpha_i = 0$ H_1 : $\exists i / \alpha_i \neq 0$

$$H_1: \exists i / \alpha_i \neq 0$$

Statistique de test : $F_{obs} = \frac{SC_A/(I-1)}{SC_B/ddl_B} = \frac{CM_A}{CM_B}$

Si
$$\forall i \ \alpha_i = \mathbf{0} \quad F_{obs} \sim F_{ddl_R}^{I-1}$$

$$F_{obs} > F_{ddl_R}^{I-1}(0.95) \Longrightarrow \text{ rejet de } H_0$$



Test global d'un effet

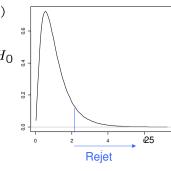
Test de l'interaction AB:

Hypothèses: $H_0: \forall (i,j), \ \alpha \beta_{ij} = 0$ $H_1: \exists (i,j) / \alpha \beta_{ij} \neq 0$

Statistique de test :
$$F_{obs} = \frac{\frac{SC_{AB}}{(I-1)(J-1)}}{\frac{SC_R}{ddl_R}} = \frac{CM_{AB}}{CM_R}$$

Si
$$\forall (i,j) \ \alpha \beta_{ij} = 0 \quad F_{obs} \sim F_{ddl_R}^{(I-1)(J-1)}$$

$$F_{obs} > F_{ddl_R}^{(I-1)(J-1)}(0.95) \Longrightarrow \text{rejet de} H_0$$
 3-



Test de conformité d'un coefficient

Hypothèses : H_0 : $\alpha_i = 0$ H_1 : $\alpha_i \neq 0$

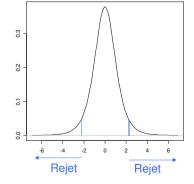
On sait que : $\mathcal{L}(\hat{\alpha}_i) = \mathcal{N}(\alpha_i, \sigma_{\hat{\alpha}_i})$ avec $\sigma_{\hat{\alpha}_i}^2 = \sigma^2 \left[(X'X)^{-1} \right]_{ii}$

d'où :
$$rac{\widehat{lpha}_i-lpha_i}{\widehat{\sigma}_{\widehat{lpha}_i}}\sim \mathcal{T}_{
u=ddl_R}$$

Statistique de test : $T_{obs} = \frac{\alpha_i}{\hat{\sigma}_{\hat{s}}}$

Si $\alpha_i = 0$, $T_{obs} \sim T_{\nu = ddl_P}$

Décision : si $|T_{obs}| > t_{\nu = ddl_R}$ (0.975) rejet de H_0



Intervalle de confiance :

$$lpha_i \in \left[\hat{lpha}_i - \hat{\sigma}_{\hat{lpha}_i} \ t_{ddl_R} (\text{0.975}) \ ; \ \hat{lpha}_i + \hat{\sigma}_{\hat{lpha}_i} \ t_{ddl_R} (\text{0.975}) \right]$$
 26

Exemple

library (FactoMineR)

AovSum(Note ~ Produit+Juge+Produit:Juge,data=donnees)

						~	_			
Prod	Juge	Note	\$Ftest	-		CM				
P1	J1	1	Prod	12.0000	1	12.0000		9	0.024	01 *
P1	J1	3	Juge	8.0000	2	4.0000		3	0.125	00
P1	J2	1	Prod:Juge	8.0000	2	4.0000		3	0.125	00
P1	J2	1	Residuals	8.0000	6	1.3333				
P1	J3	2								
		4	\$Ttest	Est	ima	te Std.	Erro	rtv	alue :	Pr(> t)
P1	J3		(Intercept				.333			0.000
P2	J1	2	Prod - P1	,				-3.0		0.024
P2	J1	2	Prod - P2					3.0		0.024
P2	J2	4								
P2	J2	6	Juge - J1					-2.1		0.078
P2	Ј3	4	Juge - J2					0.0		1.000
P2	Ј3	6	Juge - J3		1	.0 0	.471	2.1	.21	0.078
			Prod-P1 :	Juge-J1	1	.0 0	.471	2.1	.21	0.078
			Prod-P2 :	Juge-J1	-1	.0 0	.471	-2.1	.21	0.078
			Prod-P1 :	Juge-J2	-1	.0 0	.471	-2.1	.21	0.078
			Prod-P2 :	Juge-J2	1	.0 0	.471	2.1	.21	0.078
			Prod-P1 :	Juge-J3	0	0.0	.471	0.0)	1.000
			Prod-P2 :	Juge-J3	0	0.0	.471	0.0)	1.000

Extension de l'analyse de variance

Généralisation immédiate à un nombre quelconque de facteurs

Si les données sont déséquilibrées :

$$\bullet \, \widehat{\beta} = (X'X)^{-1}X'Y$$

- les SC ne s'additionnent plus
- $\hat{\alpha}_i$ et $\hat{\beta}_i$ dépendent du modèle

Modèle linéaire Contextes d'application

Modèle de régression

Sources de variabilité quantitatives

Modèle d'analyse de la variance

Sources de variabilité qualitatives

Modèle d'analyse de covariance

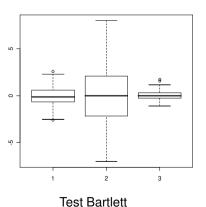
Sources de variabilité de natures différentes

Tous ces modèles sont des modèles linéaires

Analyse des **résidus**

Homoscédasticité

des **résidus** ?



résidus? 9

Normalité des

Tests Shapiro-Wilks, Kolmogorov. χ^2

27