Analyse de la variance à deux facteurs

Frédéric Bertrand[1] & Myriam Maumy[1]

[1]IRMA, Université de Strasbourg
Strasbourg, France

Master 1re Année
16-02-2012

Sommaire

Références

Ce cours s'appuie essentiellement sur

1. le livre David C. Howell, **Méthodes statistiques en sciences humaines** traduit de la sixième édition américaine aux éditions de Boeck, 2008.
2. le livre de Pierre Dagnelie, **Statistique théorique et appliquée**, Tome 2, aux éditions de Boeck, 1998.
3. le livre de Hardeo Sahai et Mohammed I. Ageel, **The Analysis of Variance : Fixed, Random and Mixed Models**, aux éditions Birkhäuser, 2000.

Sommaire

Contexte

Nous nous proposons d'analyser l'influence du temps et de trois espèces ligneuses d'arbre sur la décomposition de la masse d'une litière constituée de feuilles de Lierre.
Pour ce faire, 24 sachets d'une masse identique de feuilles de lierre ont été constitués, sachets permettant une décomposition naturelle. Puis une première série de 8 sachets, choisis au hasard, a été déposée sous un chêne, une deuxième sous un peuplier, et la dernière série sous un frêne.
Après 2, 7, 10 et 16 semaines respectivement, deux sachets sont prélevés au hasard sous chaque arbre et la masse résiduelle est déterminée pour chacun d'eux. Cette masse est exprimée en pourcentage de la masse initiale.

Les données

Les valeurs observées sont données dans le tableau suivant :

Semaine	Chêne	Peuplier	Frêne
2	85,10 87,60	85,20 84,90	84,30 85,75
7	75,90 72,85	73,00 75,70	72,80 70,80
10	71,60 66,95	74,15 71,85	67,10 64,95
16	62,10 64,30	67,25 60,25	58,75 59,00

Les écritures

Nous pouvons écrire ce tableau sous forme standard, qui est celle utilisée dans la plupart des logiciels et en particulier avec le logiciel `R`, c'est-à-dire avec trois colonnes, une pour la semaine, une pour l'espèce et une pour la masse, et 24 lignes, une pour chaque sachet.

Les données

Sachets	Semaines	Espèces	Masses
1	2	Chêne	$85,10$
2	2	Chêne	$87,60$
3	2	Peuplier	$85,20$
4	2	Peuplier	$84,90$
5	2	Frêne	$84,30$
6	2	Frêne	$85,75$
7	7	Chêne	$75,90$
8	7	Chêne	$72,85$

Les données

Sachets	Semaines	Espèces	Masses
9	7	Peuplier	73,00
10	7	Peuplier	75,70
11	7	Frêne	72,80
12	7	Frêne	70,80
13	10	Chêne	71,60
14	10	Chêne	66,95
15	10	Peuplier	74,15
16	10	Peuplier	71,85

Les données

Sachets	Semaines	Espèces	Masses
17	10	Frêne	67, 10
18	10	Frêne	64, 95
19	16	Chêne	62, 10
20	16	Chêne	64, 30
21	16	Peuplier	67, 25
22	16	Peuplier	60, 25
23	16	Frêne	58, 75
24	16	Frêne	59, 00

Le but

Nous nous proposons d'utiliser l'analyse de la variance à deux facteurs. Nous observons trois variables :

1. deux d'entre elles sont des variables contrôlées, l'espèce d'arbre, qualitative à trois modalités, et la semaine qui peut être considérée comme qualitative à quatre modalités.
2. La troisième variable est une réponse quantitative.

Donc l'analyse de la variance à deux facteurs (semaine et espèce d'arbre) croisés, avec interaction, peut convenir, entre autres méthodes d'analyse de ces données.

Sommaire

Le contexte

Dans l'étude des effets simultanés d'un facteur à I modalités et d'un facteur à J modalités sur une variable quantitative Y, supposons que Y suive des lois normales, a priori différentes dans les IJ populations disjointes déterminées par la conjonction de deux modalités des facteurs étudiés.
Supposons que, dans la population correspondant à la modalité d'ordre i du premier facteur et à la modalité d'ordre j du deuxième facteur, nous ayons :

$$\mathcal{L}(Y) = \mathcal{N}(\mu_{ij}; \sigma^2), \quad \text{pour } i = 1, \ldots, I \text{ et } j = 1, \ldots, J.$$

L'idée

Pour mettre en évidence les éventuelles différences entre le comportement de la variable Y dans les I modalités du premier facteur, dans les J modalités du deuxième facteur, ou encore dans l'interaction entre les deux facteurs, nous considérons des échantillons indépendants de même taille K de la variable Y dans chacune des IJ populations , soit au total un n-échantillon avec $n = IJK$.

Le modèle statistique

Pour la variable d'ordre k de la population d'indice (i, j), notée Y_{ijk}, nous posons :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \mathcal{E}_{ijk},$$

pour tout $i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K$, avec, pour éviter une surparamétrisation, les contraintes

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{i=1}^{I} (\alpha\beta)_{ij_0} = \sum_{j=1}^{J} (\alpha\beta)_{i_0 j} = 0,$$

pour $i_0 = 1, \ldots, I$; et $j_0 = 1, \ldots, J$.

Les hypothèses du modèle

Les variables $\mathcal{E}_{ijk}$ sont ainsi supposées suivre une loi normale $\mathcal{N}(0;\sigma^2)$.
Leurs réalisations, notées e_{ijk}, sont considérées comme les erreurs de mesure, elles sont inconnues et vérifient :

$$y_{ijk} = \mu_{ij} + e_{ijk}, \quad \text{pour } i = 1,\ldots,I; j = 1,\ldots,J; k = 1,\ldots,K.$$

Les trois tests

L'analyse de la variance à deux facteurs avec répétitions permet trois tests de Fisher. Nous testons :

- l'effet du premier facteur F_1 : Nous testons l'égalité des I paramètres α_i correspondant aux I modalités du premier facteur

$$\begin{cases} \mathcal{H}_0 : & \text{les paramètres } \alpha_i \text{ sont tous nuls} \\ \text{contre} & \\ \mathcal{H}_1 : & \text{les paramètres } \alpha_i \text{ ne sont pas tous nuls.} \end{cases}$$

Le deuxième test

Nous testons :

- l'effet du deuxième facteur F_2. Il consiste à tester l'égalité des J paramètres β_j correspondant aux J modalités du deuxième facteur

$$\begin{cases} \mathcal{H}_0 : & \text{les paramètres } \beta_j \text{ sont tous nuls} \\ \text{contre} & \\ \mathcal{H}_1 : & \text{les paramètres } \beta_j \text{ ne sont pas tous nuls.} \end{cases}$$

Le troisième test

Nous testons :

- l'effet de l'interaction entre les facteurs F_1 et F_2. Il consiste à comparer

$$\begin{cases} \mathcal{H}_0 : & \text{les } IJ \text{ paramètres } (\alpha\beta)_{ij} \text{ sont tous nuls} \\ \text{contre} & \\ \mathcal{H}_1 : & \text{les } IJ \text{ paramètres } (\alpha\beta)_{ij} \text{ ne sont pas tous nuls.} \end{cases}$$

Notations

Nous posons

$$\overline{Y} = \frac{1}{n} \sum_{i,j,k} Y_{ijk},$$

$$\overline{Y}_{ij\bullet} = \frac{1}{K} \sum_{k} Y_{ijk}, \quad \overline{Y}_{i\bullet\bullet} = \frac{1}{JK} \sum_{j,k} Y_{ijk}, \quad \overline{Y}_{\bullet j\bullet} = \frac{1}{IK} \sum_{i,k} Y_{ijk}.$$

Notations

$$SC_T = \sum_{i,j,k} (Y_{ijk} - \overline{Y})^2, \quad SC_R = \sum_{i,j,k} (Y_{ijk} - \overline{Y}_{ij\bullet})^2,$$

$$SC_\alpha = \sum_{i,j,k} (\overline{Y}_{i\bullet\bullet} - \overline{Y})^2, \quad SC_\beta = \sum_{i,j,k} (\overline{Y}_{\bullet j\bullet} - \overline{Y})^2,$$

$$SC_{\alpha\beta} = \sum_{i,j,k} (\overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y})^2.$$

L'équation de l'ANOVA

L'équation de l'analyse de la variance devient pour ce modèle :

$$SC_T = SC_R + SC_\alpha + SC_\beta + SC_{\alpha\beta}.$$

où

- la somme SC_T, la **somme totale**, mesure la somme des carrés des écarts à la moyenne globale, toutes causes confondues,

L'équation de l'ANOVA - Suite

- la somme SC_R, la **somme résiduelle**, cumule les carrés des écarts des différentes observations à la moyenne de l'échantillon dont elles font partie. Dans la somme totale elle représente la part de la dispersion due aux **fluctuations individuelles**.

L'équation de l'ANOVA - Suite

- la somme SC_α, la **somme due au premier facteur**, ou **somme entre modalités du facteur** F_α, mesure l'effet du premier facteur.
- la somme SC_β, ou **somme due au deuxième facteur**, ou **somme entre modalités du facteur** F_β, mesure l'effet du deuxième facteur.
- la somme $SC_{\alpha\beta}$ mesure l'effet de **l'interaction entre les deux facteurs**.

Propriété

Sous les différentes hypothèses $\mathcal{H}_0$ d'égalité des paramètres de la décomposition des μ_{ij}, nous pouvons préciser les lois respectives des variables précédentes. Elles suivent des lois du χ^2 :

$$\mathcal{L}_{\mathcal{H}_0}\left(\frac{1}{\sigma^2}SC_T\right) = \chi^2{}_{n-1}, \quad \mathcal{L}\left(\frac{1}{\sigma^2}SC_R\right) = \chi^2{}_{n-IJ},$$

$$\mathcal{L}_{\mathcal{H}_0}\left(\frac{1}{\sigma^2}SC_\alpha\right) = \chi^2{}_{I-1}, \quad \mathcal{L}_{\mathcal{H}_0}\left(\frac{1}{\sigma^2}SC_\beta\right) = \chi^2{}_{J-1},$$

$$\mathcal{L}_{\mathcal{H}_0}\left(\frac{1}{\sigma^2}SC_{\alpha\beta}\right) = \chi^2{}_{(I-1)(J-1)}.$$

Suite de la propriété

De plus, les variables SC_R et SC_α, SC_R et SC_β, SC_R et $SC_{\alpha\beta}$ sont indépendantes, de sorte que :

$$\mathcal{L}_{\mathcal{H}_0}\left(\frac{\dfrac{SC_\alpha}{I-1}}{\dfrac{SC_R}{IJ(K-1)}}\right) = \mathcal{F}_{(I-1),IJ(K-1)},$$

$$\mathcal{L}_{\mathcal{H}_0}\left(\frac{\dfrac{SC_\beta}{J-1}}{\dfrac{SC_R}{IJ(K-1)}}\right) = \mathcal{F}_{(J-1),IJ(K-1)},$$

Fin de la propriété

$$\mathcal{L}_{\mathcal{H}_0}\left(\frac{\dfrac{SC_{\alpha\beta}}{(I-1)(J-1)}}{\dfrac{SC_R}{IJ(K-1)}}\right) = \mathcal{F}_{(I-1)(J-1),IJ(K-1)}.$$

Les tests

Les tests sont réalisés à l'aide des valeurs numériques suivantes :

$$\overline{y} = \frac{1}{IJK}\sum_{i,j,k} y_{ijk}, \quad \overline{y}_{ij\bullet} = \frac{1}{K}\sum_{k} y_{ijk},$$

$$\overline{y}_{i\bullet\bullet} = \frac{1}{JK}\sum_{j,k} y_{ijk}, \quad \overline{y}_{\bullet j\bullet} = \frac{1}{IK}\sum_{i,k} y_{ijk}.$$

Les tests - Suite

$$
\begin{aligned}
sc_T &= \sum_{i,j,k}(y_{ijk}-\overline{y})^2 = \left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}^2\right) - IJK\overline{y}^2, \\
sc_R &= \sum_{i,j,k}(y_{ijk}-\overline{y}_{ij\bullet})^2 = \left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}^2\right) - K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{y}_{ij\bullet}^2, \\
sc_\alpha &= \sum_{i,j,k}(\overline{y}_{i\bullet\bullet}-\overline{y})^2 = JK\sum_{i=1}^{I}\overline{y}_{i\bullet\bullet}^2 - IJK\overline{y}^2,
\end{aligned}
$$

Les tests - Fin

$$
\begin{aligned}
sc_\beta &= \sum_{i,j,k}(\overline{y}_{\bullet j\bullet} - \overline{y})^2 = IK\sum_{j=1}^{J}\overline{y}_{\bullet j\bullet}^2 - IJK\overline{y}^2, \\
sc_{\alpha\beta} &= \sum_{i,j,k}(\overline{y}_{ij\bullet} - \overline{y}_{i\bullet\bullet} - \overline{y}_{\bullet j\bullet} + \overline{y})^2 \\
&= K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{y}_{ij\bullet}^2 - JK\sum_{i=1}^{I}\overline{y}_{i\bullet\bullet}^2 - IK\sum_{j=1}^{J}\overline{y}_{\bullet j\bullet}^2 + IJK\overline{y}^2.
\end{aligned}
$$

Décision

Pour un seuil $\alpha (= 5\% = 0,05$ en général), les tables de la loi de Fisher notée $\mathcal{F}$ nous fournissent pour chacun des trois tests une valeur critique c telle que $\mathbb{P}_{\mathcal{H}_0}(F < c) = 1 - \alpha$. Alors nous décidons :

$$\begin{cases} \mathcal{H}_1 \text{ est vraie si} & c \leqslant f, \\ \mathcal{H}_0 \text{ est vraie si} & f < c. \end{cases}$$

Les résultats des calculs sont généralement présentés sous la forme d'un tableau.

Tableau de l'ANOVA

Variation	SC	ddl	CM	F_{obs}	F_c
Due à F_α	sc_α	$I-1$	cm_α	$\frac{cm_\alpha}{cm_R}$	c_α
Due à F_β	sc_β	$J-1$	cm_β	$\frac{cm_\beta}{cm_R}$	c_β
Due à $F_{\alpha\beta}$	$sc_{\alpha\beta}$	$(I-1)(J-1)$	$cm_{\alpha\beta}$	$\frac{cm_{\alpha\beta}}{cm_R}$	$c_{\alpha\beta}$
Résiduelle	sc_R	$IJ(K-1)$	cm_R		
Totale	sc_T	$n-1$			

Exemple

Pour l'exemple précédent, en utilisant `R`, le tableau de l'analyse de la variance s'écrit :

	Sum Sq	Df	F	P
Semaine	1741.31	3	121.6927	$3.004e-09$
Arbre	58.08	2	6.0881	0.01495
Interaction	30.22	6	1.0559	0.43853
Résiduelle	57.24	12		
Totale	1886.84	23		

Conclusion

1. Si nous décidons $\mathcal{H}_1$, il y a **effet du premier facteur**.
2. Si nous décidons $\mathcal{H}_1$, il y a **effet du deuxième facteur**.
3. Si nous décidons $\mathcal{H}_1$, il y a **effet de l'interaction entre les deux facteurs**. Dans ce cas, pour préciser le type d'interaction mise en évidence par le test, nous pourrons comparer les moyennes $\overline{y}_{ij\bullet}$ pour les différentes valeurs de i et j.

Conclusion

Graphiquement, nous porterons en abscisse les valeurs de i (les I modalités). Pour chaque valeur de j nous relierons les valeurs de $\overline{y}_{ij\bullet}$ portées en ordonnées. L'aspect du faisceau des lignes brisées, variant ou non dans le même sens, s'interprétera facilement.

Graphique des interactions

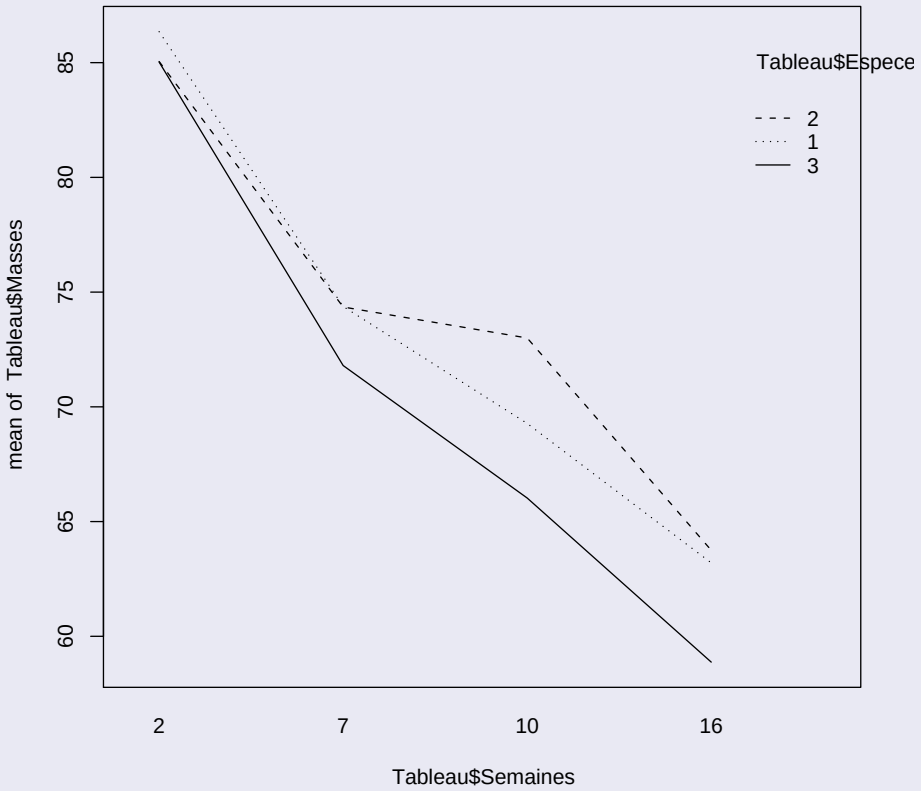

FIGURE: Représentation graphique des interactions

Graphique des interactions

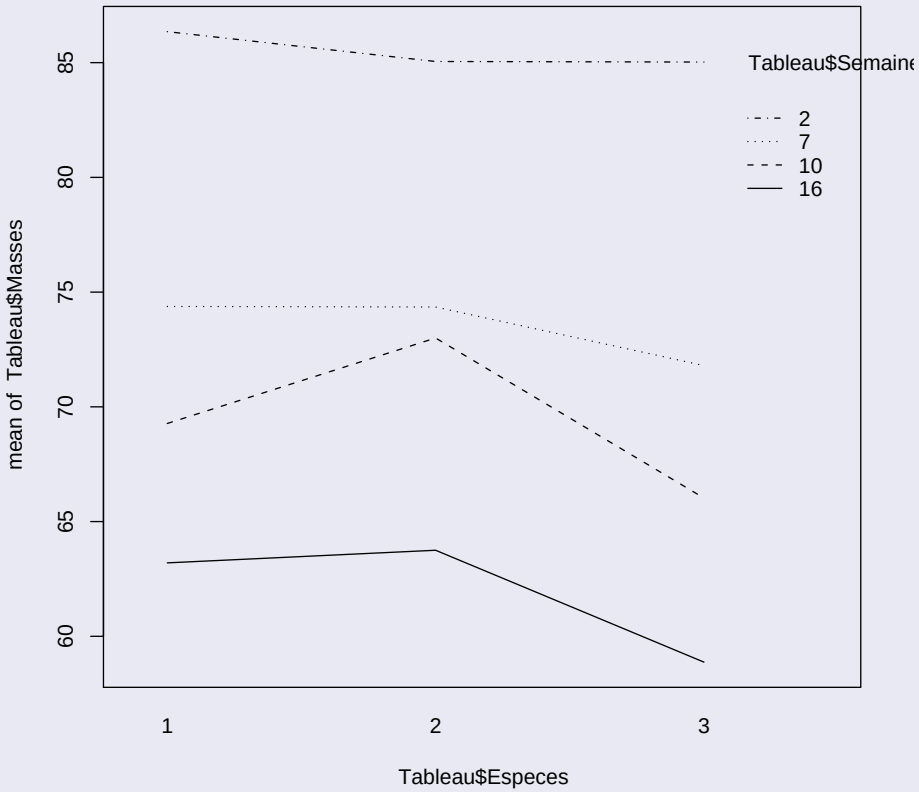

FIGURE: Représentation graphique des interactions

Sommaire

Vérification des conditions

Pour ce modèle, l'estimation des moyennes théoriques μ_{ij} se fait par les moyennes observées $\overline{y}_{ij\bullet}$ (« valeurs ajustées »). Les résidus sont alors donnés par l'expression :

$$\widehat{e}_{ijk} = y_{ijk} - \overline{y}_{ij\bullet}, \quad i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K.$$

Leur normalité et l'homogénéité des variances se vérifient par les mêmes méthodes que pour une analyse de la variance à un facteur.

Vérification des conditions sur l'exemple

- Commençons par tester la normalité des résidus.

```
> shapiro.test(residus)
Shapiro-Wilk normality test
data: residus
W = 0.9763, p-value = 0.8187
```

« L'hypothèse de normalité des résidus est acceptée (`p=0.8187`) ».

Vérification des conditions sur l'exemple - Suite et fin

- Nous allons maintenant chercher à tester l'homogénéité des variances.
  Malheureusement, dans le cas qui nous intéresse, nous ne pouvons pas tester l'homogénéité des populations parce qu'il n'y a que deux observations pour chacune d'elles et la puissance d'un tel test serait très faible. Nous verrons dans le dernier paragraphe, comment nous pouvons parer à cet inconvénient. La solution que nous proposerons n'est pas aussi performante que si nous testions l'homogénéité mais elle sera un bon indicateur pour savoir si cette condition est vérifiée.

Sommaire

Comparaisons multiples

Lorsque l'effet d'un facteur a été mis en évidence, le test de Tukey ou celui de Dunnett s'applique chaque fois que le nombre d'observations le permet, à l'aide de la même statistique. Les effectifs n_i et $n_{i'}$ sont alors ceux des classes comparées.

Comparaisons multiples sur l'exemple

Nous allons réaliser les comparaisons multiples pour le facteur « Semaines »avec R.

```
Semaines
diff lwr upr p adj
7-2 -11.966667 -15.71019 -8.2231474 0.0000033
10-2 -16.041667 -19.78519 -12.2981474
0.0000001
16-2 -23.533333 -27.27685 -19.7898141
0.0000000
10-7 -4.075000 -7.81852 -0.3314808 0.0316623
16-7 -11.566667 -15.31019 -7.8231474 0.0000047
16-10 -7.491667 -11.23519 -3.7481474 0.0003427
```

Comparaisons multiples sur l'exemple

Facteur 1 : Semaine. Nous obtenons les classes d'égalité suivantes :

Modalités du facteur	Moyennes observées	Classes d'égalité
S2	85.47500	*A*
S7	73.50833	*B*
S10	69.43333	*C*
S16	61.94167	*D*

Comparaisons multiples sur l'exemple

Nous allons réaliser les comparaisons multiples pour le facteur « Especes » avec R.

```
Especes
diff lwr upr p adj
2-1 0.73750 -2.175756 3.65075565 0.7818482
3-1 -2.86875 -5.782006 0.04450565 0.0537077
3-2 -3.60625 -6.519506 -0.69299435 0.0161088
```

Exemple

Facteur 2 : Espèces. Nous obtenons les classes d'égalité suivantes :

Modalités du facteur	Moyennes observées	Classes d'égalité
Chêne	73.30000	*A*
Peuplier	74.03750	*A*
Frêne	70.43125	*B*

Sommaire

L'idée générale

Dans le cas où nous étudions l'effet simultané de deux facteurs à, respectivement, I et J modalités et que nous disposons d'une seule observation pour chaque population, c'est à dire $K = 1$, les résultats du paragraphe précédent ne sont plus valables. Nous devons supposer que l'interaction entre les deux facteurs est nulle. Partant du même modèle, nous écrivons plus simplement :

$$Y_{ij} = \mu + \alpha_i + \beta_j + \mathcal{E}_{ij}$$

avec les contraintes $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0$.

Notations

Nous avons les notations analogues :

$$\overline{y} = \frac{1}{IJ}\sum_{i,j} y_{ij}, \quad \overline{y}_{i\bullet} = \frac{1}{J}\sum_{j} y_{ij}, \quad \overline{y}_{\bullet j} = \frac{1}{I}\sum_{i} y_{ij},$$

$$sc_T = \sum_{i,j}(y_{ij} - \overline{y})^2 = \sum_{i,j} y_{ij}^2 - IJ\overline{y}^2,$$

$$sc_R = \sum_{i,j}(y_{ij} - \overline{y}_{i\bullet} - \overline{y}_{\bullet j} + \overline{y})^2,$$

Notations

$$sc_\alpha = \sum_{i,j} (\overline{y}_{i\bullet} - \overline{y})^2 = J \sum_i \overline{y}_{i\bullet}^2 - IJ\overline{y}^2,$$

$$sc_\beta = \sum_{i,j} (\overline{y}_{\bullet j} - \overline{y})^2 = I \sum_j \overline{y}_{\bullet j}^2 - IJ\overline{y}^2.$$

Remarque importante

Remarquons que l'expression définissant, dans le cas avec répétitions, la somme des carrés associée à l'interaction, est associée ici à la somme des carrés de la résiduelle.

Tableau de l'ANOVA

Nous avons alors le tableau de l'analyse de la variance suivant :

Variation	SC	ddl	CM	F_{obs}	F_c
Due à F_α	sc_α	$I-1$	cm_α	$\frac{cm_\alpha}{cm_R}$	c_α
Due à F_β	sc_β	$J-1$	cm_β	$\frac{cm_\beta}{cm_R}$	c_β
Résiduelle	sc_R	$(I-1)(J-1)$	cm_R		
Totale	sc_T	$IJ-1$			

L'idée générale

La démarche est alors analogue à celle de l'analyse de la variance à deux facteurs avec répétitions. Notons que dans ce cas les valeurs ajustées sont données par

$$\widehat{\mu_{ij}} = \overline{y}_{i\bullet} + \overline{y}_{\bullet j} - \overline{y}$$

et les résidus par l'expression :

$$\widehat{e}_{ij} = y_{ij} - \overline{y}_{i\bullet} - \overline{y}_{\bullet j} + \overline{y}, \quad i = 1, \ldots, I; j = 1, \ldots, J.$$

Exemple

L'influence d'un traitement grossissant, à base de vitamines, est étudiée sur des animaux de races différentes. Pour cela nous disposons d'animaux de trois races, notées R_i, pour $i = 1, 2, 3$, et nous avons effectué trois traitements, notés D_j, pour $j = 1, 2, 3$, utilisant respectivement $5, 10$ et $15\mu g$ de vitamines B12 par cm^3. Le gain moyen de poids par jour est mesuré, à l'issue d'un traitement de 50 jours dans chaque cas. Un seul animal est utilisé pour chaque couple « race-traitement ».

Les données

Voici les résultats des mesures :

	R_1	R_2	R_3
D_1	$1,26$	$1,21$	$1,19$
D_2	$1,29$	$1,23$	$1,23$
D_3	$1,38$	$1,27$	$1,22$

L'objectif

Nous nous proposons d'effectuer une analyse de la variance à deux facteurs sans répétitions, il y a en effet une seule observation par « case ». Les facteurs, contrôlés, à effets fixes, sont la race et la dose, tous les deux à 3 modalités. La réponse est le gain moyen de poids.

Les données

Cet ensemble de données doit être saisi, dans un logiciel sous la forme d'un tableau empilé :

Races	Doses	Gains
R_1	D_1	$1,26$
R_1	D_2	$1,29$
R_1	D_3	$1,38$
R_2	D_1	$1,21$
R_2	D_2	$1,23$
R_2	D_3	$1,27$
R_3	D_1	$1,19$
R_3	D_2	$1,23$
R_3	D_3	$1,22$

Les hypothèses

Nous testons les hypothèses :

$$\begin{cases} \mathcal{H}_0^R : & \text{les races n'ont pas d'effet,} \\ \text{contre} & \\ \mathcal{H}_1^R : & \text{les races ont un effet} \end{cases}$$

et

$$\begin{cases} \mathcal{H}_0^D : & \text{les doses n'ont pas d'effet,} \\ \text{contre} & \\ \mathcal{H}_1^D : & \text{les doses ont un effet.} \end{cases}$$

Le tableau de l'anova

Voici le tableau de l'analyse de la variance construit par `R` :

	SumSq	Df	MeanSq	F value	Pr(>F)
Races	0.0152667	2	0.0076333	9.7447	0.02900
Doses	0.0074000	2	0.0037000	4.7234	0.08849
Res.	0.0031333	4	0.0007833		

Les résultats

Nous décidons :

1. $\mathcal{H}_1^R$ est vraie, il y a un effet de la race (`p =0.02900`)
2. $\mathcal{H}_0^D$ est vraie, il n'y a pas d'effet de la dose sur le gain de poids (`p = 0.08849`).

Vérification des hypothèses

Nous vérifions à présent les conditions d'application de l'analyse de la variance.

- Indépendance des données : Ok !
- Normalité des résidus :

```
> shapiro.test(residus)
Shapiro-Wilk normality test
data: residus
W = 0.9798, p-value = 0.9632
```

Nous décidons que l'hypothèse de normalité est vérifiée.
Nous décidons que la normalité de l'erreur théorique est acceptée.

Vérification des hypothèses

Il nous reste plus qu'à vérifier l'égalité des variances des résidus ou encore appelé l'homogénéité des variances. Remarquons tout d'abord que nous ne pouvons pas tester l'égalité des variances : en effet nous n'avons qu'une observation par « case ». Cependant, à titre indicatif, nous pouvons tester : l'égalité des variances des gains selon les races,

$$\begin{cases} \mathcal{H}_0 : & \text{les variances des races sont égales,} \\ \text{contre} & \\ \mathcal{H}_1 : & \text{les variances des races ne sont pas égales.} \end{cases}$$

Vérification des hypothèses

- Le test de Bartlett donne :

```
>
bartlett.test(residus~races,data=exemple2)
Bartlett test of homogeneity of variances
data: residus by races
Bartlett's K-squared = 3.2583, df = 2,
p-value = 0.1961
```

Nous décidons donc que l'hypothèse d'homogénéité est vérifiée. Nous décidons que les variances théoriques des gains des trois races sont égales.

Vérification des hypothèses

Nous pouvons tester aussi : l'égalité des variances des gains des doses,

$$\begin{cases} \mathcal{H}_0 : & \text{les variances des doses sont égales,} \\ \text{contre} & \\ \mathcal{H}_1 : & \text{les variances des doses ne sont pas égales.} \end{cases}$$

Vérification des hypothèses

- Le test de Bartlett donne :

```
>
bartlett.test(residus~doses,data=exemple2)
Bartlett test of homogeneity of variances
data: residus by doses
Bartlett's K-squared = 1.0819, df = 2,
p-value = 0.5822
```

Nous décidons donc que l'hypothèse d'homogénéité est vérifiée. Nous décidons que les variances théoriques des gains des trois doses sont égales.

Vérification des hypothèses

Cependant ces deux résultats ne nous garantissent pas l'égalité des 9 variances théoriques mais sont de bons indicateurs pour l'homoscédasticité.

Comparaisons multiples

Comme nous avons décidé que $\mathcal{H}_1^R$ est vraie, il y a un effet de la race, nous allons procéder à des comparaisons multiples pour analyser comment les races sont différentes par rapport au gains de poids.
Pour ce faire nous utilisons le test de Tukey au seuil de $\alpha = 5\% = 0,05$.

Script de R et sorties de R

```
> TukeyHSD(model1)
Tukey multiple comparisons of means
95% family-wise confidence level
Fit: aov(formula = gains ~ races + doses, data
= exemple2)
races
diff lwr upr p adj
2-1 -0.07333333 -0.1547783 0.008111587
0.0686703
3-1 -0.09666667 -0.1781116 -0.015221747
0.0288386
3-2 -0.02333333 -0.1047783 0.058111587
0.6040386
```

Résumé des résultats

Nous pouvons résumer ces résultats par le tableau :

Modalités du facteur	Classes d'égalité
R_1	A
R_2	B
R_3	B

Conclusions

Nous montrons ainsi que la première race est différente des deux autres, dont le gain de poids est similaire.
Comme nous avons décidé $\mathcal{H}_0^D$ est vraie, il n'y a pas d'effet de la dose sur le gain de poids, nous allons calculer le risque β a posteriori.
Pour cela, nous calculons : $\Phi = 1,575$. En reportant cette dernière valeur sur le graphique de l'abaque correspondant à $\nu_1 = J - 1 = 2$ et à $\nu_2 = (I-1)(J-1) = 4$, nous obtenons $\pi_1 = 0,45$ et $\beta = 0,55$. Ce qui signifie, que le non effet de la dose est associé à un risque de l'ordre de $0,55$ ce qui est relativement important.