République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER BISKRA

Faculté des Sciences Exactes et Sciences de la Nature et de la Vie

Département de Mathématiques

Première Année Master

Notes de Cours

Analyse de Données

Chapitre 3 : Analyse de la variance
(Séance 8)

Auteur des notes :

Dr. Sana BENAMEUR

Année universitaire : 2021-2022

3.2 Analyse de variance à deux facteurs AV(2)

Dans ce cas les données sont regroupées selon deux catégories ou facteurs : A et B. Notons $A_1, A_2, \ldots, A_p$ les p niveaux du facteur A et $B_1, B_2, \ldots, B_q$ les q niveaux du facteur B, de telle sort que chaque case contient r mesure de la variable d'intérêt Y $(y_{**1}, y_{**2}, \ldots, y_{**r})$:

	B_1	$\ldots$	B_j	$\ldots$	B_q
A_1	$y_{111}, y_{112}, \ldots, y_{11r}$	$\ldots$	$y_{1j1}, y_{1j2}, \ldots, y_{1jr}$	$\ldots$	$y_{1q1}, y_{1q2}, \ldots, y_{1qr}$
A_2	$y_{211}, y_{122}, \ldots, y_{12r}$	$\ldots$	$y_{2j1}, y_{2j2}, \ldots, y_{2jr}$	$\ldots$	$y_{2q1}, y_{2q2}, \ldots, y_{2qr}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	$y_{i11}, y_{i12}, \ldots, y_{i1r}$	$\ldots$	$y_{ij1}, y_{ij2}, \ldots, y_{ijr}$	$\ldots$	$y_{iq1}, y_{iq2}, \ldots, y_{iqr}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_p	$y_{p11}, y_{p12}, \ldots, y_{p1r}$	$\ldots$	$y_{pj1}, y_{pj2}, \ldots, y_{pjr}$	$\ldots$	$y_{pq1}, y_{pq2}, \ldots, y_{pqr}$

3.2.1 Modèle d'AV(2)

$$y_{ijk} = \underbrace{\mu}_{\text{moyenne générale}} + \underbrace{\alpha_i}_{\text{effet de } A} + \underbrace{\beta_j}_{\text{effet de } B} + \underbrace{\gamma_{ij}}_{\text{effet d'intéraction}} + \underbrace{\varepsilon_{ijk}}_{\text{effet résiduel}}$$

$$\begin{cases} \alpha_i = \mu_{i*} - \mu & \text{effet de } A \text{ au niveau } i = 1, \ldots, p \\ \beta_j = \mu_{*j} - \mu & \text{effet de } B \text{ au niveau } j = 1, \ldots, q \\ \gamma_{ij} = \mu_{ij} - \mu_{i*} - \mu_{*j} + \mu & \text{effet d'interaction entre } A \text{ et } B \\ \varepsilon_{ijk} = y_{ijk} - \mu_{ij} & \text{erreur } \ (k = 1, \ldots, r) \end{cases}$$

Les hypothèses d'intérêt dans l'AV (2) sont

$$\begin{cases} 1) \quad H_0 : \alpha_i = 0 \quad H_1 : \alpha_i \neq 0 \quad i = 1, \ldots, p \\ 2) \quad H_0 : \beta_j = 0 \quad H_1 : \beta_j \neq 0 \quad j = 1, \ldots, q \\ 3) \quad H_0 : \gamma_{ij} = 0 \quad H_1 : \gamma_{ij} \neq 0 \quad i = 1, \ldots, p, \ j = 1, \ldots, q \end{cases}$$

3.2.2 Equation d'AV(2)

Notons

$$\bar{y}_{i**} = \frac{1}{qr}\sum_{j=1}^{q}\sum_{k=1}^{r} y_{ijk} \qquad : \text{moyenne de la ligne } i$$

$$\bar{y}_{*j*} = \frac{1}{pr}\sum_{i=1}^{p}\sum_{k=1}^{r} y_{ijk} \qquad : \text{moyenne de la colonne } j$$

$$\bar{y}_{ij*} = \frac{1}{r}\sum_{k=1}^{r} y_{ijk} \qquad : \text{moyenne de la case } (i,j)$$

$$\bar{y} = \bar{y}_{***} = \frac{1}{pqr}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} y_{ijk} \qquad : \text{moyenne totale } (n = pqr)$$

L'équation d'analyse de la variance dans ce cas s'écrit

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(y_{ijk}-\bar{y})^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(y_{ijk}-\bar{y}_{ij*})^2 + qr\sum_{i=1}^{p}(\bar{y}_{i**}-\bar{y})^2 + pr\sum_{j=1}^{q}(\bar{y}_{*j*}-\bar{y})^2 + r\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{ij*}-\bar{y}_{i**}-\bar{y}_{*j*}+\bar{y})^2,$$

En d'autre terme

$$\underbrace{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} y_{ijk}^2 - n\bar{y}^2}_{SCT} = \underbrace{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(y_{ijk}-\bar{y}_{ij*})^2}_{SCR} + \underbrace{qr\sum_{i=1}^{p}\bar{y}_{i**}^2 - n\bar{y}^2}_{SCA} + \underbrace{pr\sum_{j=1}^{q}\bar{y}_{*j*}^2 - n\bar{y}^2}_{SCB} + \underbrace{r\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{ij*}-\bar{y}_{i**}-\bar{y}_{*j*}+\bar{y})^2}_{SCAB}$$

L'$AV\,(2)$ permet de tester l'absence ou l'existence des effets de A et/ou de B ou d'interaction entre A et B. Les résultats d'$AV\,(2)$ sont résumés dans le tableau suivant :

Variation	**ddl**	SC	MC	F
Fact. A	$p-1$	SCA	MCA	F_A
Fact. B	$q-1$	SCB	MCB	F_B
Fact. A,B	$(p-1)(q-1)$	$SCAB$	$MCAB$	F_{AB}
Résiduelle	$pq(r-1)$	SCR	MCR	
Totale	$n-1$	SCT		

TAB. 3.2. Tableau d'AV(2)

$$\begin{aligned}
F_A &> f_{1-\alpha}\left(p-1, pq\left(r-1\right)\right) && \Rightarrow \quad H_1 : \text{effet de } A \\
F_B &> f_{1-\alpha}\left(q-1, pq\left(r-1\right)\right) && \Rightarrow \quad H_1 : \text{effet de } B \\
F_{AB} &> f_{1-\alpha}\left(\left(p-1\right)\left(q-1\right), pq\left(r-1\right)\right) && \Rightarrow \quad H_1 : \text{effet d'interaction}
\end{aligned}$$

3.2.3 Modèle d'AV(2) sans répétitions

Supposons dans ce cas que chaque case contient ue seule observation $(r = 1)$. On dit aussi qu'une seule mesure de la variable Y pour chaque couple (A_i, B_j)

	B_1	$\dots$	B_j	$\dots$	B_q
A_1	y_{11}	$\dots$	y_{1j}	$\dots$	y_{1q}
A_2	y_{21}	$\dots$	y_{2j}	$\dots$	y_{2q}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_i	y_{i1}	$\dots$	y_{ij}	$\dots$	y_{iq}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_p	y_{p1}	$\dots$	y_{pj}	$\dots$	y_{pq}

Tableau d'AV (2) sans répétitions

Variation	ddl	SC	MC	F
Fact. A	$p-1$	SCA	MCA	F_A
Fact. B	$q-1$	SCB	MCB	F_B
Résiduelle	$(p-1)(q-1)$	SCR	MCR	
Totale	$n-1$	SCT		

Le modèle d'AV(2) est

$$y_{ij} = \underbrace{\mu}_{\text{moyenne générale}} + \underbrace{\alpha_i}_{\text{effet de } A} + \underbrace{\beta_j}_{\text{effet de } B} + \underbrace{\varepsilon_{ijk}}_{\text{effet résiduel}}$$

3.2.4 Equation d'AV(2) sans répétitions

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij}-\bar{y}\right)^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij}-\bar{y}_{i*}-\bar{y}_{*j}+\bar{y}\right)^2 + q\sum_{i=1}^{p}\left(\bar{y}_{i*}-\bar{y}\right)^2 + p\sum_{j=1}^{q}\left(\bar{y}_{*j}-\bar{y}\right)^2$$

$$\underbrace{\sum_{i=1}^{p}\sum_{j=1}^{q} y_{ij}^2 - n\bar{y}^2}_{SCT} = \underbrace{\sum_{i=1}^{p}\sum_{j=1}^{q}\left(y_{ij}-\bar{y}_{i*}-\bar{y}_{*j}+\bar{y}\right)^2}_{SCR} + \underbrace{q\sum_{i=1}^{p}\bar{y}_{i*}^2 - n\bar{y}^2}_{SCA} + \underbrace{p\sum_{j=1}^{q}\bar{y}_{*j}^2 - n\bar{y}^2}_{SCB}$$

3.2.5 Exemple d'AV(2)

Considérons les résultats d'AV(2)

Variation	ddl	SC	MC	F
Fact. A	2	105	52.5	17.5
Fact. B	4	225	112.5	37.5
Fact. A,B	8	130	65	1.67
Résiduelle	30	90	3	
Totale	44	550		

$p = 3$
$q = 5$
$r = 3$
$n = 45$

au niveau de confiance 95%, nous avons

$$F_A = 17.5 > f_{0.95}(2,30) = 3.32 \quad \Rightarrow \quad H_1 : \text{effet de } A$$
$$F_B = 37.5 > f_{0.95}(4,30) = 2.69 \quad \Rightarrow \quad H_1 : \text{effet de } B$$
$$F_{AB} = 1.67 < f_{0.95}(8,30) = 2.27 \quad \Rightarrow \quad H_0 : \text{absence d'effet d'interaction}$$

Exercice : La quantité d'oxygène consommé par deux espèces de patelle : Acmaea Scabra et Acmaea Digitalis a été analysée pour différentes conditions halines : pourcentage d'eau ; les résultats suivants ont été obtenues :

% d'eau	A.Scabra	A.Digitalis
100%	7.16 6.78 13.6 8.93 8.26	6.14 3.86 10.4 5.49 6.14
75%	5.2 5.2 7.18 6.37 13.2	4.47 9.9 5.75 11.8 4.95
50%	11.11 9.74 18.8 9.74 10.5	9.63 6.38 13.4 14.5 14.5

Analyser les résultats obtenus au cours de cette expériences au niveau : $\alpha = 95\%$.

Solution : $p = 3; q = 2; r = 5, n = pqr = 30$

	A.Scabra	A.Digitalis	$\bar{y}_{i**}$
100%	$\bar{y}_{11*} = 8.95$	$\bar{y}_{12*} = 6.41$	7.68
75%	$\bar{y}_{21*} = 7.43$	$\bar{y}_{22*} = 7.37$	7.4
50%	$\bar{y}_{31*} = 11.98$	$\bar{y}_{32*} = 11.68$	11.83
$\bar{y}_{*j*}$	9.45	8.49	$\bar{y} = 8.97$

$$
\begin{aligned}
SCT &= \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} y_{ijk}^2 - n\bar{y}^2 = 2804.37 - 30 \times (8.97)^2 = 390.54 \\
SCA &= qr\sum_{i=1}^{p} \bar{y}_{i**}^2 - n\bar{y}^2 = 10 \times \left(7.68^2 + 7.4^2 + 11.83^2\right) - 30 \times (8.97)^2 = 123.09 \\
SCB &= pr\sum_{j=1}^{q} \bar{y}_{*j*}^2 - n\bar{y}^2 = 15 \times \left(9.45^2 + 8.49^2\right) - 30 \times (8.97)^2 = 6.912 \\
SCAB &= r\sum_{i=1}^{p}\sum_{j=1}^{q} \left(\bar{y}_{ij*} - \bar{y}_{i**} - \bar{y}_{*j*} + \bar{y}\right)^2 \\
&= 5 \times (0.6241 + 0.6241 + 0.2025 + 0.2025 + 0.1089 + 0.1089) \\
&= 5 \times 1.871 \\
&= 9.355 \\
SCR &= SCT - SCA - SCB - SCAB = 251.185 \\
MCA &= SCA/(p-1) = 61.545; MCB = SCB/(q-1) = 6.912 \\
MCAB &= SCAB/(p-1)(q-1) = 4.68; MCR = SCR/pq(r-1) = 10.47
\end{aligned}
$$

Le tabeau d'AV(2) est donc

Variation	**ddl**	SC	MC	F
Fact. A	2	123.09	61.545	5.88
Fact. B	1	6.912	6.91	0.66
Fact. A,B	2	9.355	4.68	
Résiduelle	24	251.185	10.47	
Totale	29	390.54		

$$
\begin{aligned}
F_A &= 5.88 > f_{0.95}(2,24) = 3.40 \Rightarrow H_1 \text{ effet de } A \text{ (\% d'eau)} \\
F_B &= 0.66 < f_{0.95}(1,24) = 4.26 \Rightarrow H_0 \text{ absence d'effet de } B \text{ (espèces de patelle)} \\
F_{AB} &= 0.45 < f_{0.95}(2,24) = 3.40 \Rightarrow H_0 \text{ absence d'effet d'interaction}
\end{aligned}
$$