Chapitre IV

Analyse du plan $S \times 0$

El Methni M.

IV-1. *O* est un (seul) facteur à effets fixes

IV-1-1 Généralités

O est un facteur à *r* modalités o_1, o_2, … o_r (Occasions). S est le facteur sujet (à effets aléatoires) dont les modalités sont les *n* sujets s_1, s_2, … s_n. Pour chaque modalité du facteur *O* on « utilise » les *n* sujets. On a donc des mesures y_{si} répétées pour chaque sujet *s* dans l'occasion (modalité) O_i.
En psychologie et en méthodologie on parle de plan intra, en statistique on parle de plan à mesures répétées.
On dispose donc de $N=n \times r$ mesures y_{si} que l'on présente sous la forme d'un tableau :

	Facteur *O* (Occasions)					
Sujets	o_1	o_2	…	o_i	…	o_r
s_1	y_{11}	y_{12}	…	y_{1i}	…	y_{1r}
s_2	y_{21}	y_{22}	…	y_{2i}	…	y_{2r}
…	…	…	…		…	…
s_s	y_{s1}	y_{s2}	…	y_{si}	…	y_{sr}
…	…	…	…	…	…	…
s_n	y_{n1}	y_{n2}		y_{ni}		y_{nr}

Remarque : Un premier avantage d'un tel plan par rapport au plan $S<O>$ est une « économie » du nombre de sujets. D'autre part avec un tel plan on peut diminuer l'erreur expérimentale (résidu). En effet le contrôle du facteur sujet permet de séparer son effet des effets résiduels. Mais le résidu sera confondu avec l'interaction entre *S* et *O*.

IV-1-2 Modèle univarié :

Les données peuvent être d'une part, considérées comme les $N=n \times r$ observations de la variable réponse *Y* dans les $N=n \times r$ conditions expérimentales décrites par le croisement du facteur sujet et du facteur occasion. Elles peuvent d'autre part, être regardées comme les *n* observations d'un vecteur de *r* variables réponses. Ces deux points de vue conduisent à l'élaboration de deux modèles, le modèle mixte univarié dans le premier cas et le modèle multivarié dans le second (voir condition de validation IV-1-5).
Nous étudierons le modèle mixte univarié dans lequel on considère que les modalités du facteur sujet ont été obtenues par échantillonnage, le facteur sujet est donc un facteur aléatoire. Le facteur occasion est lui un facteur à effets fixes.
Chaque donnée y_{si} correspond alors à l'observation d'une variable aléatoire réelle Y_{si} décrite par le modèle suivant : $Y_{si} = \mu_i + \pi_s + \varepsilon_{si}$
Où :

- les μ_i sont des constantes mesurant les effets fixes des modalités *i* ($i=1, 2, \ldots, r$) du facteur *O*
- les π_s sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2_\pi)$ mesurant les effets aléatoires des modalités *s* ($s=1, 2, \ldots, n$) du facteur *S*
- les résidus ε_{si} sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$. On suppose en plus que les π_s et les ε_{si} sont indépendantes.

On peut réécrire l'effet de la modalité *i* sous la forme : $\mu_i = \mu + \alpha_i$ on a donc : $Y_{si} = \mu + \alpha_i + \pi_s + \varepsilon_{si}$
Où :

- μ constante s'interprétant comme un niveau général de réponse

- les α_i sont des constantes mesurant les effets fixes des modalités i et vérifiant : $\sum_i \alpha_i = 0$ (contrainte d'identifiabilité)

Conséquence : D'après ce modèle les Y_{si} sont des variables aléatoires de loi $\mathcal{N}(\mu+\alpha_i, \sigma^2_\pi+\sigma^2)$.
L'effet du facteur fixe se traduit sur la moyenne de la variable réponse tandis que l'effet du facteur aléatoire se traduit sur la variance de la variable réponse.
Pour ce modèle, l'interaction entre le facteur sujet et le facteur occasion est confondue avec le résidu car pour chaque modalité de l'un et de l'autre on ne dispose que d'une seule observation.

IV-1-3 Décomposition de la variation :
On décompose la variation totale en variation inter-sujets (due au facteur sujet) et variation intra-sujets (due à l'effet du facteur O). $\mathbf{SCT}_{\text{obs}} = \mathbf{SC}_{\text{inter-sujets}} + \mathbf{SC}_{\text{intra-sujets}}$
La variation inter-sujets est due au facteur sujet et la variation intra-sujets est due d'une part à l'effet du facteur O et d'autre part à l'effet des autres facteurs non contrôlés

$$\mathbf{SC}_{\text{intra-sujets}} = \mathbf{SCO}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$$

Conclusion : Dans le cas du plan complet $S\times O$, la variation totale se décompose de façon additive en : $\mathbf{SCT}_{\text{obs}} = \mathbf{SCS}_{\text{obs}} + \mathbf{SCO}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$
De même on a la décomposition des degrés de liberté : nr-1 = (n-1) + (r-1) + (n-1)(r-1)
Les différentes sommes de carrées se calculent par :

$$\mathbf{SCT}_{\text{obs}} = \sum_{s=1}^{n}\sum_{i=1}^{r}(y_{si} - \bar{y})^2 = nr \times \text{variance totale}$$

$$\mathbf{SCS}_{\text{inter-sujet}} = \mathbf{SCS}_{\text{obs}} = r\sum_{s=1}^{n}(\bar{y}_{s0} - \bar{y})^2 = nr \times \text{variance des moyennes des sujets}$$

$$\mathbf{SCO}_{\text{obs}} = n\sum_{i=1}^{r}(\bar{y}_{0i} - \bar{y})^2 = nr \times \text{variance des moyennes par occasion}$$

SCR est calculé par : $\mathbf{SCR}_{\text{obs}} = \mathbf{SCT}_{\text{obs}} - \mathbf{SCS}_{\text{obs}} - \mathbf{SCO}_{\text{obs}}$

En pratique on peut effectuer les calculs en utilisant le tableau suivant :

	Facteur O (Occasions)							
Sujets	o_1	o_2	…	o_i	…	o_r	$\bar{y}_{s0}$	$(\bar{y}_{s0} - \bar{y})^2$
s_1	y_{11}	y_{12}	…	y_{1i}	…	y_{1r}	$\bar{y}_{10}$	$(\bar{y}_{10} - \bar{y})^2$
s_2	y_{21}	y_{22}	…	y_{2i}	…	y_{2r}	$\bar{y}_{20}$	$(\bar{y}_{20} - \bar{y})^2$
…	…	…	…		…	…	…	…
s_s	y_{s1}	y_{s2}	…	y_{si}	…	y_{sr}	$\bar{y}_{s0}$	$(\bar{y}_{s0} - \bar{y})^2$
…	…	…	…	…	…	…	…	…
s_n	y_{n1}	y_{n2}		y_{ni}		y_{nr}	$\bar{y}_{n0}$	$(\bar{y}_{n0} - \bar{y})^2$
$\bar{y}_{0i}$	$\bar{y}_{01}$	$\bar{y}_{02}$	…	$\bar{y}_{0i}$	…	$\bar{y}_{0r}$	$\bar{y}_{00} = \bar{y}$	$r\sum_{s=1}^{n}(\bar{y}_{s0} - \bar{y})^2$
$(\bar{y}_{0i} - \bar{y})^2$	$(\bar{y}_{01} - \bar{y})^2$	$(\bar{y}_{02} - \bar{y})^2$		$(\bar{y}_{0i} - \bar{y})^2$		$(\bar{y}_{0r} - \bar{y})^2$	$n\sum_{i=1}^{r}(\bar{y}_{0i} - \bar{y})^2$	

IV-1-4 Statistiques des tests :
On montre le théorème suivant :

Théorème : sous les hypothèses du modèle on a • E(**SCO**) = $(r - 1)\sigma^2 + \sum n\alpha_i^2$ • E(**SCS**) = $(n - 1)(\sigma^2 + r\sigma^2_\pi)$ • E(**SCR**) = $(n - 1)(r - 1)\sigma^2$

En considérant les moyennes des carrés : $\mathbf{MCO} = \dfrac{\mathbf{SCO}}{r-1}$ $\quad \mathbf{MCR} = \dfrac{\mathbf{SCR}}{(r-1)(n-1)}$

On peut réécrire le théorème précédent sous la forme :

Théorème : sous les hypothèses du modèle on a • $E(\mathbf{MCO}) = \sigma^2 + \dfrac{n}{r-1}\sum \alpha_i^2$ • $E(\mathbf{MCS}) = \sigma^2 + \dfrac{r}{n-1}\sigma_\pi^2$ • $E(\mathbf{MCR}) = \sigma^2$

De nouveau on constate que **MCR** est un estimateur sans biais de la variance σ^2 et que **MCO** est un également un estimateur de la variance σ^2 augmentée d'une valeur positive traduisant l'effet du facteur. On considérera donc la statistique $\mathbf{F}_O = \dfrac{\mathbf{MCO}}{\mathbf{MCR}}$ pour effectuer le test suivant :

<u>Test de l'effet du facteur *O* sur la variable *Y* :</u>

hypothèse nulle $\mathbf{H_0}$: *O* n'a pas d'effet $\Leftrightarrow (\forall i) \quad \alpha_i = 0$ (tous les α_i sont nuls) contre
l'hypothèse alternative : $\mathbf{H_1}$: *O* a un effet $\Leftrightarrow (\exists i) \quad \alpha_i \neq 0$ (l'un au moins des α_i est non nul).

Sous l'hypothèse nulle **MCO** et **MCR** sont deux estimateurs indépendants de σ^2 , leur rapport $\mathbf{F_O}$ est distribué selon une loi de Fischer à $\nu_1 = r - 1$ et $\nu_2 = (n - 1)(r - 1)$ ddl.

Le test est alors défini au seuil de signification α par la règle de décision suivante :

si $\mathbf{F_O}_{\text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(\mathbf{F_O} \geq \lambda_\alpha)$.

Les résultats sont présentés dans le tableau d'analyse de la variance suivant :

Source de variation	SC	ddl	MC	F_O
Inter *S*	$\mathbf{SCS}_{obs}$	$n - 1$	$\mathbf{MCS}_{obs}$	
Intra *S*	$\mathbf{SC}_{\text{intra-sujets}}$	$n(r – 1)$		
O	$\mathbf{SCO}_{obs}$	$r - 1$	$\mathbf{MCO}_{obs}$	$\mathbf{F_O}_{\text{ obs}}$
R	$\mathbf{SCR}_{obs}$	$(n - 1)(r - 1)$	$\mathbf{MCR}_{obs}$	
Total	$\mathbf{SCT}_{obs}$	$rn - 1 = N - 1$		

IV-1-5 Condition de validation :

Les données peuvent être d'une part, considérées comme les $N=n\times r$ observations de la variable réponse *Y* dans les $N=n\times r$ conditions expérimentales décrites par le croisement du facteur sujet et du facteur occasion. Elles peuvent d'autre part, être regardées comme les *n* observations d'un vecteur de *r* variables réponses. Ces deux points de vue conduisent à l'élaboration de deux modèles, le <u>modèle mixte univarié</u> dans le premier cas et le <u>modèle multivarié</u> dans le second.

Le fait de mesurer plusieurs fois la variable réponse sur le même sujet introduit des corrélations entre les observations faites sur ce même sujet. Dans le cas du modèle mixte univarié on montre que :

$$\operatorname{cov}(Y_{si},Y_{s'i'}) = \begin{cases} 0 \text{ si } s \neq s' \\ \sigma_\pi^2 \text{ si } s = s' \; et \; i \neq i' \\ \sigma_\pi^2 + \sigma^2 \text{ si } s = s' \; et \; i = i' \end{cases}$$

$$\text{et } cor(Y_{si},Y_{s'i'}) = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma^2}$$

Dans le cas du modèle multivarié on considère que les données sont les réalisations de *r* vecteurs aléatoires de dimension *n*. $Y_s=(Y_{s1}, Y_{s2}, \ldots, Y_{sr})$ indépendants et de même loi normale caractérisés par : $E(Y_{si})=\mu_i$ Var$(Y_{si})=\sigma^2_i$ cov$(Y_{si}, Y_{si'})$ = cov$_{ii'}$

Le modèle mixte univarié est donc un cas particulier du modèle multivarié se caratérisant par un <u>matrice de variance-covariance</u> Σ de la forme :

$$\Sigma = \begin{array}{c|cccccc} & O_1 & O_2 & \cdots & O_i & \cdots & O_r \\ \hline O_1 & \sigma_\pi^2+\sigma^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2 & \cdots & \sigma_\pi^2 \\ O_2 & \sigma_\pi^2 & \sigma_\pi^2+\sigma^2 & \cdots & \sigma_\pi^2 & \cdots & \sigma_\pi^2 \\ \cdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ O_i & \sigma_\pi^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2+\sigma^2 & \cdots & \sigma_\pi^2 \\ \cdots & \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ O_r & \sigma_\pi^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2 & \cdots & \sigma_\pi^2+\sigma^2 \end{array}$$

Une telle matrice où tous les termes sont positifs et où tous les termes diagonaux sont égaux et tous les autres sont égaux est dite <u>matrice circulaire</u>, et cette propriété est appelée la <u>symétrie composée</u> de la matrice.

Sur la diagonale principale de la matrice des variances-covariances Σ se trouvent les variances et les éléments hors diagonale sont des covariances.

L'hypothèse d'homogénéité des variances (*homoscédasticité*) se traduit par l'égalité des éléments de la diagonale de la matrice Σ, Les éléments en dehors de la diagonale principale de la matrice Σ doivent être de même ordre.

De même on peut considérer la matrice des <u>corrélations</u> ρ :

$$\rho = \begin{array}{c|cccccc} & O_1 & O_2 & \ldots & O_i & \ldots & O_r \\ \hline O_1 & 1 & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} \\ O_2 & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & 1 & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ O_i & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & 1 & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ O_r & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & 1 \end{array}$$

De ce fait, dans le cas du choix de modèle mixte univarié, en calculant les variances-covariances nous obtenons $\hat{\Sigma}$ l'estimation de la matrice des variances-covariances (et aussi l'estimation de la matrice des corrélations).

Si les éléments de la diagonale de la matrice des variances-covariances $\hat{\Sigma}$ sont de même ordre, alors la condition de l'homoscédasticité (homogénéité des variances) est vérifiée.

Si les éléments hors diagonale de la matrice des variances-covariance (ou des corrélations) sont de même ordre que les variances, alors le choix du modèle mixte univarié est justifié. C'est-à-dire que la condition de la circularité (de sphéricité) est acquise.

IV-1-6 Exemple :

Pendant ¼ d'heure on compte le nombre d'actions exercées par chacun des 7 rats sur un levier. Et ceci dans trois conditions de renforcement :
1^{ere} condition : on présente au rat des aliments très appréciés.
2eme condition : on présente au rat des aliments moyennement appréciés.
3eme condition : on présente au rat des aliments peu appréciés.
On obtient les résultats donnés par le tableau ci-contre.

	Facteur O				
Sujets	A	B	C	moyenne	$(\bar{y}_{s0}-\bar{y})^2$
s_1	8	6	2	$\bar{y}_{10}$=5,333	1,416
s_2	6	5	1	$\bar{y}_{20}$=4	0,020
s_3	7	5	0	$\bar{y}_{30}$=4	0,020
s_4	9	3	3	$\bar{y}_{40}$=5	0,734
s_5	5	4	1	$\bar{y}_{50}$=3,333	0,656
s_6	7	5	2	$\bar{y}_{60}$=4,667	0,274
s_7	6	2	0	$\bar{y}_{70}$=2,667	2,178
moyenne	$\bar{y}_{01}$=6,857	$\bar{y}_{02}$=4,286	$\bar{y}_{03}$=1,286	$\bar{y}=T/N$ =4,143	5,298×3=15,90
$(\bar{y}_{0i}-\bar{y})^2$	7,365	0,020	8,162	15,547×7=108,857	

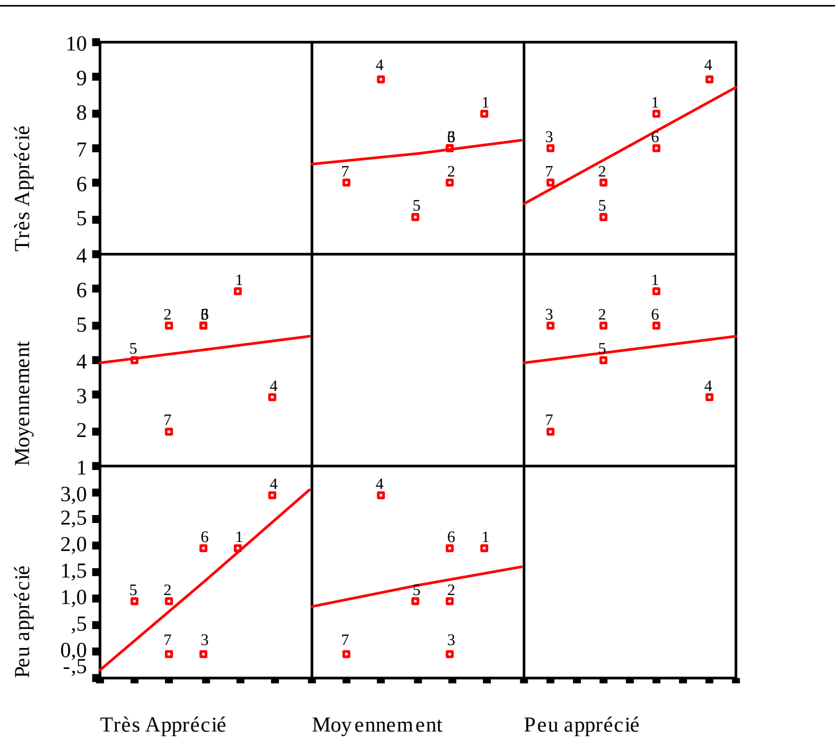
SCT=138,5714 **SCO**=108,857
SCS=15,90 **SCR**=**SCT- SCO- SCS**=13,8

$$\mathbf{MCO}=\frac{\mathbf{SCO}}{r-1}=\frac{108{,}9}{2}=54{,}45 \quad \mathbf{MCR}=\frac{\mathbf{SCR}}{(n-1)(r-1)}=\frac{13{,}8}{6\times 2}=1{,}15 \quad \mathbf{F}_{\text{obs}}=\frac{\mathbf{MCO}}{\mathbf{MCR}}=\frac{54{,}45}{1{,}15}=47{,}34$$

Les résultats sont présentés dans le tableau d'analyse de la variance suivant :

Source de variation	**SC**	ddl	**MC**	$\mathbf{F_O}$
Inter S	$\mathbf{SCS}_{obs}$=15,90	n – 1=6	$\mathbf{MCS}_{obs}$=2,65	
Intra S	$\mathbf{SC}_{\text{intra-sujets}}$	$n(r$ – 1)=14		
O	$\mathbf{SCO}_{obs}$=108,857	r – 1=2	$\mathbf{MCO}_{obs}$=54,45	$\mathbf{F_O}_{\,obs}$=47,34
R	$\mathbf{SCR}_{obs}$=13,8	(n - 1)(r - 1)=12	$\mathbf{MCR}_{obs}$=1,15	
Total	$\mathbf{SCT}_{obs}$=138,5714	rn - 1 =N – 1=20		

Pour vérifier la condition de la symétrie composée, nous allons établir la matrice des variances-covariances et aussi la matrice des corrélations pour les colonnes du tableau des observations.



Matrice des variances-covariances

$$\begin{array}{c} \\ O_1 \\ O_2 \\ O_3 \end{array} \begin{array}{c} \begin{array}{ccc} O_1 & O_2 & O_3 \end{array} \\ \begin{bmatrix} 1,551 & & \\ 0,184 & 1,633 & \\ 0,898 & 0,204 & 1,061 \end{bmatrix} \end{array}$$

Matrice des corrélations

$$\begin{array}{c} \\ O_1 \\ O_2 \\ O_3 \end{array} \begin{array}{c} \begin{array}{ccc} O_1 & O_2 & O_3 \end{array} \\ \begin{bmatrix} 1 & & \\ 0,115 & 1 & \\ 0,700 & 0,155 & 1 \end{bmatrix} \end{array}$$

On constate que les estimations des variances ne sont pas très différentes. L'hypothèse d'homogénéité des variances (homoscédasticité) peut être considérée comme acquise.

En ce qui concerne les covariances ou les corrélations, on constate une corrélation relativement forte entre les résultats des conditions 1 et 3 ($r_{13} = 0,7$).

Dans la condition 3 (aliments peu appréciés), les rats n'ont pas trop d'actions. Mais, les rats très actifs dans la condition 1 (aliments très appréciés) sont aussi actifs dans la condition 3, alors que les résultats de la condition 2 ne sont pas corrélés ni avec la condition 1 ni avec la condition 2.

Ces résultats suggèrent une réflexion supplémentaire sur ces données. Peut être faut-il examiner un modèle multivarié et effectuer une *MANOVA*.

El Methni M.

IV-1. *O* est un (seul) facteur à effets fixes

IV-1-1 Généralités

O est un facteur à *r* modalités o_1, o_2, … o_r (Occasions). S est le facteur sujet (à effets aléatoires) dont les modalités sont les *n* sujets s_1, s_2, … s_n. Pour chaque modalité du facteur *O* on « utilise » les *n* sujets. On a donc des mesures y_{si} répétées pour chaque sujet *s* dans l'occasion (modalité) O_i.
En psychologie et en méthodologie on parle de plan intra, en statistique on parle de plan à mesures répétées.
On dispose donc de $N=n\times r$ mesures y_{si} que l'on présente sous la forme d'un tableau :

	Facteur *O* (Occasions)					
Sujets	o_1	o_2	…	o_i	…	o_r
s_1	y_{11}	y_{12}	…	y_{1i}	…	y_{1r}
s_2	y_{21}	y_{22}	…	y_{2i}	…	y_{2r}
…	…	…	…		…	…
s_s	y_{s1}	y_{s2}	…	y_{si}	…	y_{sr}
…	…	…	…	…	…	…
s_n	y_{n1}	y_{n2}		y_{ni}		y_{nr}

Remarque : Un premier avantage d'un tel plan par rapport au plan *S*<*O*> est une « économie » du nombre de sujets. D'autre part avec un tel plan on peut diminuer l'erreur expérimentale (résidu). En effet le contrôle du facteur sujet permet de séparer son effet des effets résiduels. Mais le résidu sera confondu avec l'interaction entre *S* et *O*.

IV-1-2 Modèle univarié :

Les données peuvent être d'une part, considérées comme les $N=n\times r$ observations de la variable réponse *Y* dans les $N=n\times r$ conditions expérimentales décrites par le croisement du facteur sujet et du facteur occasion. Elles peuvent d'autre part, être regardées comme les *n* observations d'un vecteur de *r* variables réponses. Ces deux points de vue conduisent à l'élaboration de deux modèles, le modèle mixte univarié dans le premier cas et le modèle multivarié dans le second (voir condition de validation IV-1-5).
Nous étudierons le modèle mixte univarié dans lequel on considère que les modalités du facteur sujet ont été obtenues par échantillonnage, le facteur sujet est donc un facteur aléatoire. Le facteur occasion est lui un facteur à effets fixes.
Chaque donnée y_{si} correspond alors à l'observation d'une variable aléatoire réelle Y_{si} décrite par le modèle suivant : $Y_{si} = \mu_i + \pi_s + \varepsilon_{si}$
Où :

- les μ_i sont des constantes mesurant les effets fixes des modalités *i* ($i=1, 2, …, r$) du facteur *O*
- les π_s sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2_\pi)$ mesurant les effets aléatoires des modalités *s* ($s=1, 2, …, n$) du facteur *S*
- les résidus ε_{si} sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, \sigma^2)$. On suppose en plus que les π_s et les ε_{si} sont indépendantes.

On peut réécrire l'effet de la modalité *i* sous la forme : $\mu_i = \mu + \alpha_i$ on a donc : $Y_{si} = \mu + \alpha_i + \pi_s + \varepsilon_{si}$
Où :

- μ constante s'interprétant comme un niveau général de réponse
- les α_i sont des constantes mesurant les effets fixes des modalités *i* et vérifiant :
  $\sum_i \alpha_i = 0$ (contrainte d'identifiabilité)

Conséquence : D’après ce modèle les Y_{si} sont des variables aléatoires de loi $\mathcal{N}(\mu+\alpha_i, \sigma^2_\pi+\sigma^2)$.
L’effet du facteur fixe se traduit sur la moyenne de la variable réponse tandis que l’effet du facteur aléatoire se traduit sur la variance de la variable réponse.
Pour ce modèle, l’interaction entre le facteur sujet et le facteur occasion est confondue avec le résidu car pour chaque modalité de l’un et de l’autre on ne dispose que d’une seule observation.

IV-1-3 Décomposition de la variation :
On décompose la variation totale en variation inter-sujets (due au facteur sujet) et variation intra-sujets (due à l’effet du facteur O). $\mathbf{SCT}_{\text{obs}} = \mathbf{SC}_{\text{inter-sujets}} + \mathbf{SC}_{\text{intra-sujets}}$
La variation inter-sujets est due au facteur sujet et la variation intra-sujets est due d’une part à l’effet du facteur O et d’autre part à l’effet des autres facteurs non contrôlés

$$\mathbf{SC}_{\text{intra-sujets}} = \mathbf{SCO}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$$

Conclusion : Dans le cas du plan complet $S{\times}O$, la variation totale se décompose de façon additive en : $\mathbf{SCT}_{\text{obs}} = \mathbf{SCS}_{\text{obs}} + \mathbf{SCO}_{\text{obs}} + \mathbf{SCR}_{\text{obs}}$
De même on a la décomposition des degrés de liberté : nr-1 = (n-1) + (r-1) + (n-1)(r-1)
Les différentes sommes de carrées se calculent par :

$$\mathbf{SCT}_{\text{obs}} = \sum_{s=1}^{n}\sum_{i=1}^{r}(y_{si}-\bar{y})^2 = nr \times \text{variance totale}$$

$$\mathbf{SCS}_{\text{inter-sujet}} = \mathbf{SCS}_{\text{obs}} = r\sum_{s=1}^{n}(\bar{y}_{s0}-\bar{y})^2 = nr \times \text{variance des moyennes des sujets}$$

$$\mathbf{SCO}_{\text{obs}} = n\sum_{i=1}^{r}(\bar{y}_{0i}-\bar{y})^2 = nr \times \text{variance des moyennes par occasion}$$

SCR est calculé par : $\mathbf{SCR}_{\text{obs}} = \mathbf{SCT}_{\text{obs}} - \mathbf{SCS}_{\text{obs}} - \mathbf{SCO}_{\text{obs}}$

En pratique on peut effectuer les calculs en utilisant le tableau suivant :

	Facteur O (Occasions)							
Sujets	o_1	o_2	…	o_i	…	o_r	$\bar{y}_{s0}$	$(\bar{y}_{s0}-\bar{y})^2$
s_1	y_{11}	y_{12}	…	y_{1i}	…	y_{1r}	$\bar{y}_{10}$	$(\bar{y}_{10}-\bar{y})^2$
s_2	y_{21}	y_{22}	…	y_{2i}	…	y_{2r}	$\bar{y}_{20}$	$(\bar{y}_{20}-\bar{y})^2$
…	…	…	…		…	…	…	…
s_s	y_{s1}	y_{s2}	…	y_{si}	…	y_{sr}	$\bar{y}_{s0}$	$(\bar{y}_{s0}-\bar{y})^2$
…	…	…	…	…	…	…	…	…
s_n	y_{n1}	y_{n2}		y_{ni}		y_{nr}	$\bar{y}_{n0}$	$(\bar{y}_{n0}-\bar{y})^2$
$\bar{y}_{0i}$	$\bar{y}_{01}$	$\bar{y}_{02}$	…	$\bar{y}_{0i}$	…	$\bar{y}_{0r}$	$\bar{y}_{00}=\bar{y}$	$r\sum_{s=1}^{n}(\bar{y}_{s0}-\bar{y})^2$
$(\bar{y}_{0i}-\bar{y})^2$	$(\bar{y}_{01}-\bar{y})^2$	$(\bar{y}_{02}-\bar{y})^2$		$(\bar{y}_{0i}-\bar{y})^2$		$(\bar{y}_{0r}-\bar{y})^2$	$n\sum_{i=1}^{r}(\bar{y}_{0i}-\bar{y})^2$	

IV-1-4 Statistiques des tests :
On montre le théorème suivant :

Théorème : sous les hypothèses du modèle on a
- E(**SCO**) = $(r - 1)\sigma^2 + \sum n\alpha_i^2$
- E(**SCS**) = $(n - 1)(\sigma^2 + r\sigma^2_\pi)$

- $E(\mathbf{SCR}) = (n - 1)(r - 1)\sigma^2$

En considérant les moyennes des carrés : $\mathbf{MCO} = \frac{\mathbf{SCO}}{r-1}$ $\mathbf{MCR} = \frac{\mathbf{SCR}}{(r-1)(n-1)}$

On peut réécrire le théorème précédent sous la forme :

Théorème : sous les hypothèses du modèle on a

- $E(\mathbf{MCO}) = \sigma^2 + \frac{n}{r-1}\sum \alpha_i^2$
- $E(\mathbf{MCS}) = \sigma^2 + \frac{r}{n-1}\sigma_\pi^2$
- $E(\mathbf{MCR}) = \sigma^2$

De nouveau on constate que **MCR** est un estimateur sans biais de la variance σ^2 et que **MCO** est un également un estimateur de la variance σ^2 augmentée d'une valeur positive traduisant l'effet du facteur. On considérera donc la statistique $\mathbf{F}_O = \frac{\mathbf{MCO}}{\mathbf{MCR}}$ pour effectuer le test suivant :

Test de l'effet du facteur *O* sur la variable *Y* :

hypothèse nulle $\mathbf{H_0}$: *O* n'a pas d'effet $\Leftrightarrow (\forall i)\ \ \alpha_i = 0$ (tous les α_i sont nuls) contre
l'hypothèse alternative : $\mathbf{H_1}$: *O* a un effet $\Leftrightarrow (\exists i)\ \ \alpha_i \neq 0$ (l'un au moins des α_i est non nul).

Sous l'hypothèse nulle **MCO** et **MCR** sont deux estimateurs indépendants de σ^2 , leur rapport $\mathbf{F_O}$ est distribué selon une loi de Fischer à $\nu_1 = r - 1$ et $\nu_2 = (n - 1)(r - 1)$ ddl.

Le test est alors défini au seuil de signification α par la règle de décision suivante :

si $\mathbf{F_O}_{\text{ obs}} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle

où λ_α est donné par l'équation : $\alpha = P(\mathbf{F_O} \geq \lambda_\alpha)$.

Les résultats sont présentés dans le tableau d'analyse de la variance suivant :

Source de variation	**SC**	ddl	**MC**	$\mathbf{F_O}$
Inter *S*	$\mathbf{SCS}_{\text{obs}}$	$n - 1$	$\mathbf{MCS}_{\text{obs}}$	
Intra *S*	$\mathbf{SC}_{\text{intra-sujets}}$	$n(r-1)$		
O	$\mathbf{SCO}_{\text{obs}}$	$r - 1$	$\mathbf{MCO}_{\text{obs}}$	$\mathbf{F_O}_{\text{ obs}}$
R	$\mathbf{SCR}_{\text{obs}}$	$(n - 1)(r - 1)$	$\mathbf{MCR}_{\text{obs}}$	
Total	$\mathbf{SCT}_{\text{obs}}$	$rn - 1 = N - 1$		

IV-1-5 Condition de validation :

Les données peuvent être d'une part, considérées comme les $N=n\times r$ observations de la variable réponse *Y* dans les $N=n\times r$ conditions expérimentales décrites par le croisement du facteur sujet et du facteur occasion. Elles peuvent d'autre part, être regardées comme les *n* observations d'un vecteur de *r* variables réponses. Ces deux points de vue conduisent à l'élaboration de deux modèles, le modèle mixte univarié dans le premier cas et le modèle multivarié dans le second.

Le fait de mesurer plusieurs fois la variable réponse sur le même sujet introduit des corrélations entre les observations faites sur ce même sujet. Dans le cas du modèle mixte univarié on montre que :

$$\mathrm{cov}(Y_{si}, Y_{s'i'}) = \begin{cases} 0 \text{ si } s \neq s' \\ \sigma_\pi^2 \text{ si } s = s' \; et \; i \neq i' \\ \sigma_\pi^2 + \sigma^2 \text{ si } s = s' \; et \; i = i' \end{cases}$$

$$\text{et } cor(Y_{si}, Y_{s'i'}) = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma^2}$$

Dans le cas du modèle multivarié on considère que les données sont les réalisations de *r* vecteurs aléatoires de dimension *n*. $Y_s=(Y_{s1}, Y_{s2}, \ldots, Y_{sr})$ indépendants et de même loi normale caractérisés par : $E(Y_{si})=\mu_i$ $\mathrm{Var}(Y_{si})=\sigma^2_i$ $\mathrm{cov}(Y_{si}, Y_{si'}) = \mathrm{cov}_{ii'}$

Le modèle mixte univarié est donc un cas particulier du modèle multivarié se caratérisant par un matrice de variance-covariance Σ de la forme :

$$\Sigma = \begin{array}{c} \\ O_1 \\ O_2 \\ \cdots \\ O_i \\ \cdots \\ O_r \end{array} \begin{array}{c} \begin{array}{cccccc} O_1 & O_2 & \cdots & O_i & \cdots & O_r \end{array} \\ \begin{bmatrix} \sigma_\pi^2+\sigma^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2 & \cdots & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2+\sigma^2 & \cdots & \sigma_\pi^2 & \cdots & \sigma_\pi^2 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \sigma_\pi^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2+\sigma^2 & \cdots & \sigma_\pi^2 \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \sigma_\pi^2 & \sigma_\pi^2 & \cdots & \sigma_\pi^2 & \cdots & \sigma_\pi^2+\sigma^2 \end{bmatrix} \end{array}$$

Une telle matrice où tous les termes sont positifs et où tous les termes diagonaux sont égaux et tous les autres sont égaux est dite matrice circulaire, et cette propriété est appelée la symétrie composée de la matrice.

Sur la diagonale principale de la matrice des variances-covariances Σ se trouvent les variances et les éléments hors diagonale sont des covariances.

L'hypothèse d'homogénéité des variances (*homoscédasticité*) se traduit par l'égalité des éléments de la diagonale de la matrice Σ, Les éléments en dehors de la diagonale principale de la matrice Σ doivent être de même ordre.

De même on peut considérer la matrice des corrélations ρ :

$$\rho = \begin{array}{c} \\ O_1 \\ O_2 \\ \ldots \\ O_i \\ \ldots \\ O_r \end{array} \begin{array}{c} \begin{array}{cccccc} O_1 & O_2 & \ldots & O_i & \ldots & O_r \end{array} \\ \begin{bmatrix} 1 & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} \\ \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & 1 & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & 1 & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & \frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma^2} & \ldots & 1 \end{bmatrix} \end{array}$$

De ce fait, dans le cas du choix de modèle mixte univarié, en calculant les variances-covariances nous obtenons $\hat{\Sigma}$ l'estimation de la matrice des variances-covariances (et aussi l'estimation de la matrice des corrélations).

Si les éléments de la diagonale de la matrice des variances-covariances $\hat{\Sigma}$ sont de même ordre, alors la condition de l'homoscédasticité (homogénéité des variances) est vérifiée.

Si les éléments hors diagonale de la matrice des variances-covariance (ou des corrélations) sont de même ordre que les variances, alors le choix du modèle mixte univarié est justifié. C'est-à-dire que la condition de la circularité (de sphéricité) est acquise.

IV-1-6 Exemple :

Pendant ¼ d'heure on compte le nombre d'actions exercées par chacun des 7 rats sur un levier. Et ceci dans trois conditions de renforcement :
1^{ere} condition : on présente au rat des aliments très appréciés.
2eme condition : on présente au rat des aliments moyennement appréciés.
3eme condition : on présente au rat des aliments peu appréciés.
On obtient les résultats donnés par le tableau ci-contre.

	Facteur O				
Sujets	A	B	C	moyenne	$(\bar{y}_{s0}-\bar{y})^2$
s_1	8	6	2	$\bar{y}_{10}$=5,333	1,416
s_2	6	5	1	$\bar{y}_{20}$=4	0,020
s_3	7	5	0	$\bar{y}_{30}$=4	0,020
s_4	9	3	3	$\bar{y}_{40}$=5	0,734
s_5	5	4	1	$\bar{y}_{50}$=3,333	0,656
s_6	7	5	2	$\bar{y}_{60}$=4,667	0,274
s_7	6	2	0	$\bar{y}_{70}$=2,667	2,178
moyenne	$\bar{y}_{01}$=6,857	$\bar{y}_{02}$=4,286	$\bar{y}_{03}$=1,286	$\bar{y}=T/N$ =4,143	5,298×3=15,90
$(\bar{y}_{0i}-\bar{y})^2$	7,365	0,020	8,162	15,547×7=108,857	

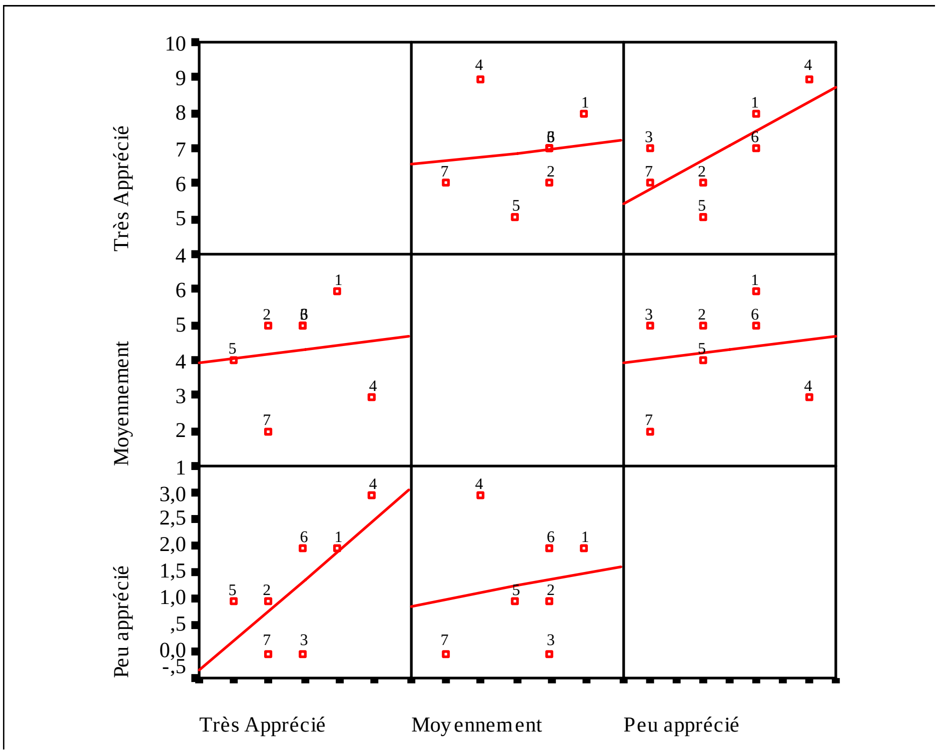
SCT=138,5714 **SCO**=108,857
SCS=15,90 **SCR**=**SCT**- **SCO**- **SCS**=13,8

$$\mathbf{MCO}=\frac{\mathbf{SCO}}{r-1}=\frac{108{,}9}{2}=54{,}45 \quad \mathbf{MCR}=\frac{\mathbf{SCR}}{(n-1)(r-1)}=\frac{13{,}8}{6\times 2}=1{,}15 \quad \mathbf{F}_{\text{obs}}=\frac{\mathbf{MCO}}{\mathbf{MCR}}=\frac{54{,}45}{1{,}15}=47{,}34$$

Les résultats sont présentés dans le tableau d'analyse de la variance suivant :

Source de variation	**SC**	ddl	**MC**	$\mathbf{F_O}$
Inter S	$\mathbf{SCS}_{\text{obs}}$=15,90	$n-1$=6	$\mathbf{MCS}_{\text{obs}}$=2,65	
Intra S	$\mathbf{SC}_{\text{intra-sujets}}$	$n(r-1)$=14		
O	$\mathbf{SCO}_{\text{obs}}$=108,857	$r-1$=2	$\mathbf{MCO}_{\text{obs}}$=54,45	$\mathbf{F_O}_{\text{ obs}}$=47,34
R	$\mathbf{SCR}_{\text{obs}}$=13,8	$(n-1)(r-1)$=12	$\mathbf{MCR}_{\text{obs}}$=1,15	
Total	$\mathbf{SCT}_{\text{obs}}$=138,5714	$rn-1=N-1$=20		

Pour vérifier la condition de la symétrie composée, nous allons établir la matrice des variances-covariances et aussi la matrice des corrélations pour les colonnes du tableau des observations.



Matrice des variances-covariances

$$\begin{array}{cc} & \begin{array}{ccc} O_1 & O_2 & O_3 \end{array} \\ \begin{array}{c} O_1 \\ O_2 \\ O_3 \end{array} & \begin{bmatrix} 1,551 & & \\ 0,184 & 1,633 & \\ 0,898 & 0,204 & 1,061 \end{bmatrix} \end{array}$$

Matrice des corrélations

$$\begin{array}{cc} & \begin{array}{ccc} O_1 & O_2 & O_3 \end{array} \\ \begin{array}{c} O_1 \\ O_2 \\ O_3 \end{array} & \begin{bmatrix} 1 & & \\ 0,115 & 1 & \\ 0,700 & 0,155 & 1 \end{bmatrix} \end{array}$$

On constate que les estimations des variances ne sont pas très différentes. L'hypothèse d'homogénéité des variances (homoscédasticité) peut être considérée comme acquise.

En ce qui concerne les covariances ou les corrélations, on constate une corrélation relativement forte entre les résultats des conditions 1 et 3 ($r_{13} = 0,7$).

Dans la condition 3 (aliments peu appréciés), les rats n'ont pas trop d'actions. Mais, les rats très actifs dans la condition 1 (aliments très appréciés) sont aussi actifs dans la condition 3, alors que les résultats de la condition 2 ne sont pas corrélés ni avec la condition 1 ni avec la condition 2.

Ces résultats suggèrent une réflexion supplémentaire sur ces données. Peut être faut-il examiner un modèle multivarié et effectuer une *MANOVA*.