14 ANOVA à deux voies

La généralisation de l'ANOVA à une voie à des plans d'expérience plus complexes est assez intuitive. L'exemple suivant utilise deux facteurs.

Exemple. Le loyer moyen dans une grande ville française, en fonction de deux facteurs : *Date de construction* A et *Nombre de pièces* B est donné par

		$A = 1$	2	3	4
		< 1981	1981-1970	1991-2001	> 2001
$B = 1$	1 pièce	509	503	521	795
2	2 pièces	596	661	814	1138
3	3 pièces	684	791	1071	1503
4	4 pièces	808	960	1259	1741
5	5 pièces	1075	1216	1604	2131

L'analyse d'un tel jeu de données à pour objectif d'expliquer et de quantifier l'influence des deux facteurs sur la variable réponse (le loyer). Le plan d'expérience employé est complet, en ce sens que, pour chaque combinaison des deux facteurs, on dispose d'une observation.

14.1 Position du problème

On veut mesurer maintenant le rôle conjoint de deux facteurs A et B sur la variable dépendante (réponse). Trois effets sont à mesurer :

- effet de A,
- effet de B,
- interaction entre A et B.

Les deux premiers seront les effets principaux.

14.1.1 Description des données

La population est notée P, X est la variable d'intérêt de moyenne globale μ. On étudie le rôle de deux facteurs A et B, le facteur A ayant p modalités $(A_1, \cdots A_p)$, le facteur B ayant q modalités $(B_1, \cdots B_q)$.

- Les facteurs A et B définissent $p \times q$ sous population P_{ij}, sur laquelle la sous-variable X_{ij} de X prend les observations $x_{i,j,k}$ (pour $k \leq n_{ij}$) et a pour moyenne μ_{ij}.

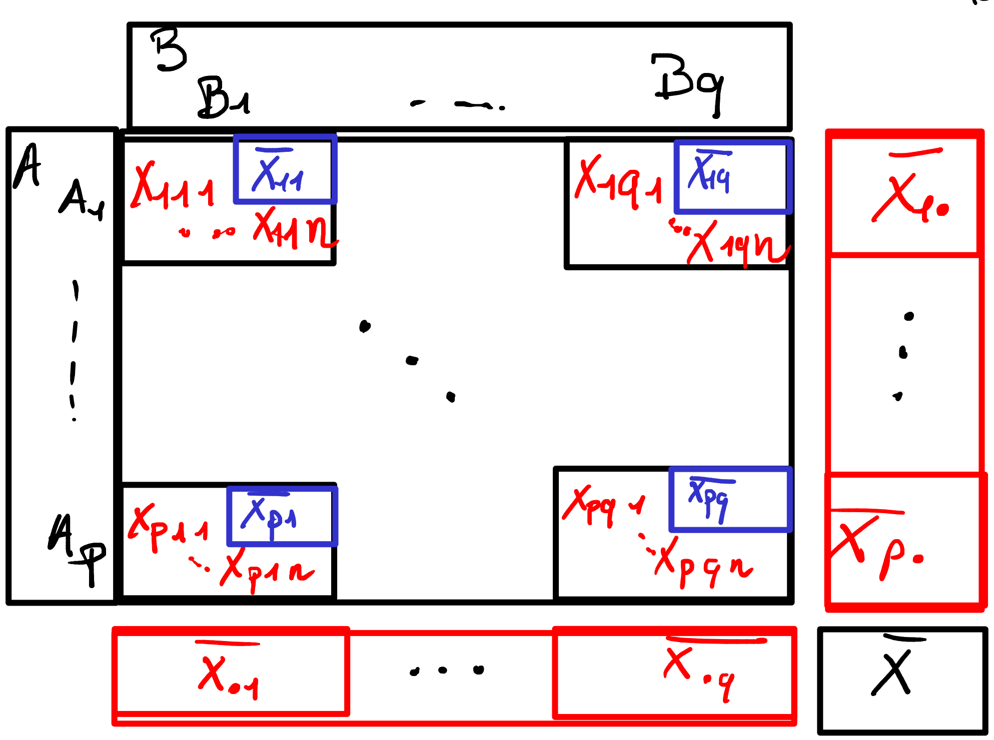
- On note $P_{i.}$ les individus correspondants à $A = A_i$, la sous variable $X_{i.}$ de X est observée par la concaténation sur j des $x_{i,j,k}$. La variable $X_{i.}$ a pour moyenne $\mu_{i.}$ et effectif $n_{i.} = \sum_{j=1}^{q} n_{ij}$.
- On note $P_{.j}$ les individus correspondants à $B = B_j$, la sous variable $X_{.j}$ de X est observée par la concaténation sur j des $x_{i,j,k}$. La variable X_j a pour moyenne $\mu_{.j}$ et effectif $n_{j.} = \sum_{i=1}^{p} n_{ij}$.

On suppose que dans chaque sous-population P_{ij}, les observations $x_{i,j,k}$ forme un échantillon E_{ij}.

on prend le même nombre d'observations pour tout couple de modalités de facteurs (A, B)
$\Rightarrow$ effectif n par couple de modalité.

Note. Pour simplifier l'exposé, dans tout ce qui suit, on considère que le plan d'expériences est équilibré, $card(E_{ij}) = n_{ij} = n$. Cela n'est pas gênant, en effet, en passant par un plan équilibré, on améliore la robustesse du test. D'autre part, le traitement par R saura s'occuper des différences d'effectif s'il y en a.

$$\bar{X}_{ij} = \sum_{k=1}^{n} \frac{X_{ijk}}{n}$$



$$\bar{X}_{i\cdot} = \frac{1}{nq}\sum_j n\bar{X}_{ij} = \frac{1}{q}\sum_j \bar{X}_{ij}$$

$$\bar{X}_{\cdot j} = \frac{1}{p}\sum_i \bar{X}_{ij}$$

$$\bar{X} = \frac{1}{npq}\sum_j np\,\bar{X}_{\cdot j} = \frac{1}{npq}\sum_i nq\,\bar{X}_{i\cdot}$$

$$= \frac{1}{q}\sum_j \bar{X}_{\cdot j} = \frac{1}{p}\sum_i \bar{X}_{i\cdot}$$

14.2 Tableau de données

Facile à lire mais encombrant, pratique pour les calculs manuels.

aspiration carburant	atmo	turbo
diesel	52 56 (...) 58	68 65 (...) 67
essence	48 49 (...) 67	102 145 (...) 130

$$x_{2,2,2} = puissance_{essence,turbo,2}$$

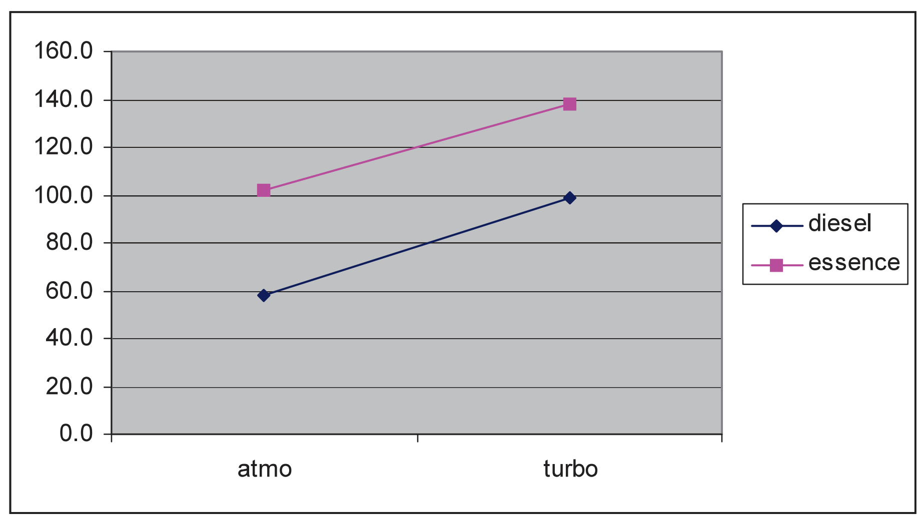
Dans cet exemple, que nous allons filer dans la suite du cours, le tableau n'est pas exhaustif, nous n'avons mis que les trois premières valeurs de chaque groupe.

14.3 Tableau des moyennes

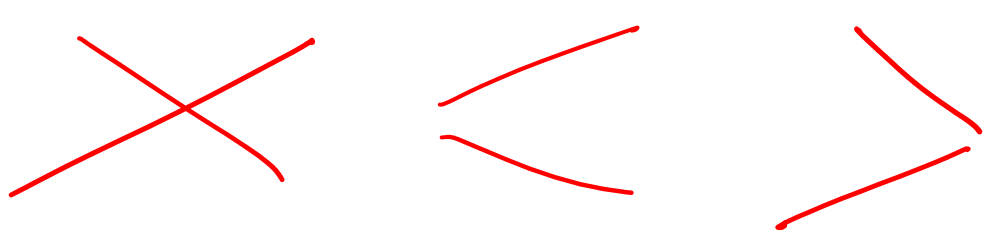
Moyenne puissance	Aspiration atmo	turbo	Total
diesel	$\bar{X}_{11} = 58.1$	$\bar{X}_{12} = 98.6$	$\bar{X}_{1.} = 78.35$
essence	$\bar{X}_{21} = 101.6$	$\bar{X}_{22} = 138.4$	$\bar{X}_{2.} = 120$
Total	$\bar{X}_{.1} = 79.85$	$\bar{X}_{.2} = 118.5$	$\bar{X} = 112.1$

14.4 Graphe des interactions

Ce graphe permet de distinguer les interactions lorsque les lignes se croisent.



droites parallèles (ou presque) ⇒ pas d'effet de l'interaction



⇒ effet de l'interaction

Dans la suite, on considère l'effet de l'interaction.

<u>Modèle</u> : $X_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ $\quad\quad \varepsilon_{ijk} \sim \mathcal{N}(0,\sigma)$ i.i.d.

$\mu_{ij} = \mu + \underbrace{(\mu_{i\cdot} - \mu)}_{\text{effet de } A} + \underbrace{(\mu_{\cdot j} - \mu)}_{\text{effet de } B} + \underbrace{(\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu)}_{\text{effet de l'interact}^{\circ} \text{ de } A,B}$

14.5 Hypothèses statistiques

Ce sont les mêmes que pour l'ANOVA à 1 facteur : normalité de la variable dépendante, indépendance des observations inter et intra groupes, variance homogène dans les groupes.

$X_{ijk} \sim \mathcal{N}(\mu_{ij}, \sigma)$ indépendantes

14.6 Hypothèses soumises au test

Il y en a trois :

$$\begin{cases} H_0 \;:\; \mu_{i.} = \mu, \quad \forall i & \text{absence d'effet de } A \\ H_0 \;:\; \mu_{.j} = \mu, \quad \forall j & \text{absence d'effet de } B \\ H_0 \;:\; \mu_{ij} = \mu, \quad \forall i,j & \text{absence d'effet de l'interaction.} \end{cases}$$

Équation de la variance.

$$x_{i,j,k} - \bar{X} = \underbrace{(\bar{X}_{i.} - \bar{X}) + (\bar{X}_{.j} - \bar{X})}_{\substack{\text{Effet des facteurs} \\ \text{principaux}}} + \underbrace{(\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X})}_{\substack{\text{Effet de} \\ \text{l'interacton}}} + \underbrace{(x_{i,j,k} - \bar{X}_{ij})}_{\substack{\text{Erreur} \\ \text{résiduelle}}}.$$

ε_{ijk}

À partir de laquelle on extrait l'équation d'ANOVA

$$\underbrace{SCT}_{\substack{\text{Variablité} \\ \text{totale}}} = \underbrace{SCM_A + SCM_B + SCM_{AB}}_{\substack{\text{Variabilité} \\ \text{expliquée}}} + \underbrace{SCE}_{\substack{\text{Variabilité} \\ \text{résiduelle}}}.$$

$SCT = \sum (x_{ijk} - \bar{X})^2$ ... $SCM_{AB} = SCT - SCM_A - SCM_B - SCE$

$SCM_A = nq \sum_{i=1}^{p} (\bar{X}_{i.} - \bar{X})^2$

$SCM_B = np \sum_{j=1}^{q} (\bar{X}_{.j} - \bar{X})^2$

$SCE = \sum (x_{ijk} - \bar{X}_{ij})^2$

Carrés moyens.

$$CM. = \frac{SC.}{DDL}.$$

$$CMT = \frac{SCT}{\underbrace{npq}-1} \longrightarrow \sum_{ijk}(X_{ijk}-\bar{X})^2$$

effectif total.

$$CME = \frac{SCE}{\underbrace{npq}_{\text{effectif total}} - \underbrace{pq}_{\text{groupes}}} \longrightarrow \sum_{ijk}\left(X_{ijk}-\bar{X}_{ij}\right)^2$$

connus

⇒ −1 ddl par groupe.

$$CME = \frac{SCE}{(n-1)\,pq}$$

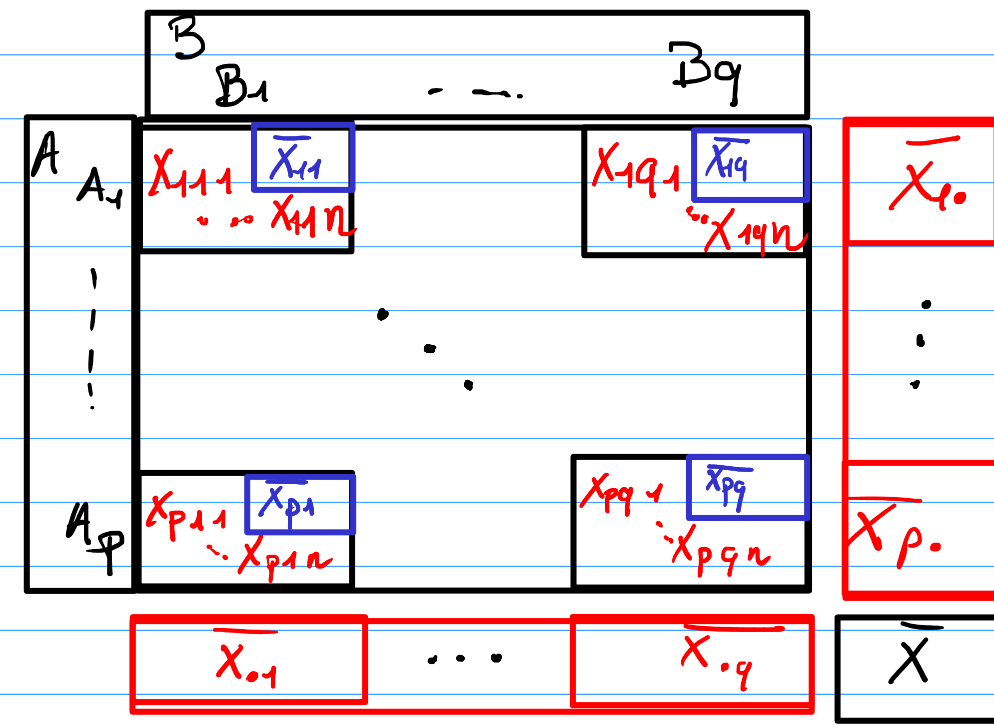
$$CMM_A = \frac{SCM_A}{p-1} \longrightarrow nq\sum_{i=1}^{p}(\bar{X}_{i.}-\bar{X})^2$$

p moyennes et $\bar{X}$ fixé.

$$CMM_B = \frac{SCM_B}{q-1}$$

$$CMM_{AB} = \frac{SCM_{AB}}{(p-1)(q-1)} \longrightarrow \sum_i(\bar{X}_{ij} - \dots)^2$$

$\bar{X}_{i.}$ et $\bar{X}_{.j}$ ⇒ 1 ddl de moins par ligne et colonne



On calcule les carrés moyens

$$CMT = \tfrac{SCT}{pqn-1},$$

$$CMM_A = \frac{SCM_A}{p-1}, \qquad CMM_B = \tfrac{SCM_B}{q-1}, \qquad CMM_{AB} = \frac{SCM_{AB}}{(p-1)(q-1)},$$

$$CME = \tfrac{SCE}{pq(n-1)}.$$

$$\mathcal{H}_{0A} = \text{"} \mu_{1\cdot} = \mu_{2\cdot} = \dots = \mu_{p\cdot} = \mu \text{"} \iff \text{"} \underbrace{\sum_i (\mu_{i\cdot} - \mu)^2}_{SCM_A} = 0 \text{"}$$

$$\mathcal{H}_{0B} = \text{"} \mu_{\cdot 1} = \mu_{\cdot 2} = \dots = \mu_{\cdot q} = \mu \text{"} \iff \text{"} \underbrace{\sum_j (\mu_{\cdot j} - \mu)^2}_{SCM_B} = 0 \text{"}$$

$$\mathcal{H}_{0AB} = \text{"} \mu_{ij} = \mu \text{"} \iff \text{"} \underbrace{\sum_i (\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu)^2}_{SCM_{AB}} = 0 \text{"}$$

Voici les rapports de carrés moyens pour construire les statistiques de Fisher utilisées pour mettre à jour les effets (principaux et interactions).

- Effet de A :

$$F_A = \frac{CMM_A}{CME}. = \frac{SCM_A/(p-1)}{SCE/(n-1)pq}$$

- Effet de B :

$$F_B = \frac{CMM_B}{CME}. = \frac{SCM_B/(q-1)}{SCE/(n-1)pq}$$

- Effet de l'interaction de A et de B :

$$F_{AB} = \frac{CMM_{AB}}{CME}.$$

Ces quantités suivent une loi de Fischer, les degrés de libertés sont lus dans les dénominateurs des carrés moyens associés.

$$F_{AB} = \frac{SCM_{AB}/(p-1)(q-1)}{SCE/(n-1)pq}.$$

14.8 ANOVA à deux facteurs sous R

Puissances des véhicules en fonction du type de carburant (`Fuel-type`) et le mode d'alimentation (`aspiration`).

diesel	gas
84.450	106.396

std	turbo
99.81105	124.43243

```
#Données pour ANOVA à 2 facteurs
autos.2 <- read.xlsx("autos_anova.xlsx",header=T,sheetIndex=2)
print(summary(autos.2))

#moyennes conditionnelles
#vs. fuel.type
print(tapply(autos.2$horsepower,list(autos.2$fuel.type),mean))

#vs. aspiration
print(tapply(autos.2$horsepower,list(autos.2$aspiration),mean))

#vs. fuel.type * aspiration
print(tapply(autos.2$horsepower,list(autos.2$fuel.type,autos.2$aspirati
on),mean))

#ANOVA à 2 facteurs
fit2 <- aov(horsepower ~ (fuel.type + aspiration) + fuel.type*aspiration,
data = autos.2)
print(summary(fit2))
```

	std	turbo
diesel	58.14286	98.61538
gas	101.62271	138.41667

avec interaction

On obtient le tableau d'ANOVA à deux facteurs suivant

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fuel.type	1	8693	8693	6.373	0.0124 *
aspiration	1	35678	35678	26.156	7.32e-07 ***
fuel.type:aspiration	1	51	51	0.037	0.8475
Residuals	201	274179	1364		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

// le carburant a un effet significatif sur la puissance.

// la carburation a un effet significatif sur la puissance

// l'interaction des facteurs n'a pas d'effet significatif.