Démarche utilisée : nous comparons la dispersion des moyennes (8, 10, 12) à la dispersion à l'intérieur de chaque groupe.

Comparer a moyennes sur des groupes indépendants

Plan d'expérience : $\mathcal{S} < \mathcal{A}_a >$

Une variable $\mathcal{A}$, de modalités $A_1, A_2, \ldots, A_a$ définit a groupes indépendants.

Variable dépendante X mesurée sur chaque sujet.
x_{ij} : valeur observée sur le i-ème sujet du groupe j.

Problème : La variable X a-t-elle la même moyenne dans chacune des sous-populations dont les groupes sont issus ?

H_0 : $\mu_1 = \mu_2 = \ldots = \mu_a$
H_1 : Les moyennes ne sont pas toutes égales.

Construction de la statistique de test :

Notations :
n_1, n_2, ..., n_a : effectifs des groupes.
N : effectif total
$T_{\cdot 1}$, ..., $T_{\cdot a}$: sommes des observations pour chacun des groupes.
$T_{\cdot\cdot}$ ou T_G : somme de toutes les observations.

Somme des carrés totale ou variation totale :

$$SC_T = \sum_{i,j} x_{ij}^2 - \frac{T_G^2}{N}$$

Elle se décompose en une variation “intra-groupes” et une variation “inter-groupes” :

$SC_T = SC_{inter} + SC_{intra}$ avec :

$$SC_{inter} = \sum_{j=1}^{a} \frac{T_{\cdot j}^2}{n_j} - \frac{T_G^2}{N}$$

$$SC_{intra} = \sum_{i,j} x_{ij}^2 - \sum_{j=1}^{a} \frac{T_{\cdot j}^2}{n_j}$$

Carrés moyens :

$$CM_{inter} = \frac{SC_{inter}}{a-1} \quad ; \quad CM_{intra} = \frac{SC_{intra}}{N-a}$$

Statistique de test : $F = \dfrac{CM_{inter}}{CM_{intra}}$

F suit une loi de Fisher à $(a-1)$ et $(N-a)$ ddl.

Présentation des résultats

Source de variation	SC	*ddl*	CM	F
$\mathcal{A}$ (inter-groupes)	SC_{inter}	$a-1$	CM_{inter}	F_{obs}
Résiduelle (intra-gr.)	SC_{intra}	$N-a$	CM_{intra}	
Total	SC_T	$N-1$		

Organisation des calculs

$i \quad j$	1	2	3	Total
1	x_{11}			
...	...	...	...	
$T_{\cdot j}$	$T_{\cdot 1}$			T_G
$T_{\cdot j}^2$				
n_j				N
$\frac{T_{\cdot j}^2}{n_j}$				
$\sum x_{ij}^2$				

Exemple :

15 sujets évaluent 3 couvertures de magazine. Sont-elles équivalentes ?

	C1	C2	C3
	14	16	14
	6	14	16
	12	8	14
	10	8	14
	8	14	12
$\overline{x}_i$	10	12	14

Calculs

$i \quad j$	1	2	3	Total
1	$x_{11} = 14$	16	14	
2	$x_{21} = 6$	14	16	
...	...	...	...	
$T_{\cdot j}$	50	60	70	$T_G = 180$
$T_{\cdot j}^2$	2500	3600	4900	
n_j	5	5	5	$N = 15$
$\frac{T_{\cdot j}^2}{n_j}$	500	720	980	2200
$\sum x_{ij}^2$	540	776	988	2304

$$SC_{inter} = \sum_{j=1}^{a} \frac{T_{\cdot j}^2}{n_j} - \frac{T_G^2}{N} = 2200 - \frac{180^2}{15} = 40$$

$$SC_{intra} = \sum_{i,j} x_{ij}^2 - \sum_{j=1}^{a} \frac{T_{\cdot j}^2}{n_j} = 2304 - 2200 = 104$$

$$SC_T = \sum_{i,j} x_{ij}^2 - \frac{T_G^2}{N} = 144$$

$$CM_{inter} = \frac{SC_{inter}}{a-1} = 20 \; ; \; CM_{intra} = \frac{SC_{intra}}{N-a} = 8.67$$

$$F_{obs} = \frac{CM_{inter}}{CM_{intra}} = 2.31$$

F suit une loi de Fisher avec $ddl_1 = a - 1 = 2$ et $ddl_2 = N - a = 12$.

Résultats

Source	Somme carrés	ddl	Carré Moyen	F
$\mathcal{C}$	40	2	20	2.31
Résid.	104	12	8.67	
Total	144	14		

Pour α=5%, $F_{crit} = 3.88$: H_0 est acceptée

Remarques

– SC_{inter} : c'est la somme des carrés (totale) que l'on obtiendrait si toutes les observations d'un groupe étaient égales à la moyenne de ce groupe.
CM_{inter} : variance corrigée de cet ensemble de données.

– SC_{intra} : c'est la somme des carrés (totale) que l'on obtiendrait en "décalant" chaque observation de façon à avoir la même moyenne dans chaque groupe.
CM_{intra} : "moyenne pondérée" des trois variances corrigées ainsi obtenues.

– Hypothèses "a priori" :
distribution normale de X dans chacun des groupes
Egalité des variances dans les trois populations.

– Si 2 groupes, équivaut à un T de Student. $F = T^2$

Pour les deux situations proposées en introduction :

Situation 1

Analysis of Variance Table

Response: x1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	80.000	40.000	17.008	1.659e-05 ***
Residuals	27	63.500	2.352		

Situation 2

Analysis of Variance Table

Response: x2

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	80.00	40.00	2.7136	0.08436 .
Residuals	27	398.00	14.74		

Vocabulaire des plans d'expérience

Variable dépendante

– On formule une hypothèse : "telle variable a tel effet sur le comportement des sujets"

– On choisit une **variable dépendante**

Définir une variable mesurable (numérique) caractérisant le comportement du sujet.

Qualités d'une bonne variable dépendante : *pertinence, sensibilité.*

Variables indépendantes ou facteurs

– Recherche des **variables indépendantes** ou **facteurs de variation**

Indépendant : indépendant du sujet, manipulé ou contrôlé par l'expérimentateur

Causes susceptibles d'entraîner une variation de la variable dépendante.

Les facteurs sont des variables nominales ou ordinales. Les valeurs prises par un facteur sont ses *modalités* ou *niveaux*.

Les *Facteurs principaux* sont ceux dont on désire étudier l'effet. Ils sont aussi appelés *facteurs d'intérêt*.

Les *Facteurs secondaires* sont les autres causes susceptibles d'influer sur le comportement des sujets. Deux manières de les prendre en compte :

– Contrôle
– Neutralisation.

Facteur systématique ou fixe : l'ensemble des modalités possibles est fini (et petit). Toutes les modalités sont présentes dans l'expérience.

Facteur aléatoire : l'ensemble des modalités est grand (infini). On choisit alors (par tirage au sort) un ensemble de modalités.

Le facteur sujet : généralement, c'est un facteur aléatoire et secondaire. Il est souvent assimilé à une incertitude sur une mesure.

Facteur étiquette : par exemple, le sexe, ou le milieu socio-culturel.

Groupe contrôle

Souvent, il existe un niveau particulier ou “état nul” de la VI. Un groupe de sujets soumis à ce niveau de la VI constitue un *groupe contrôle*.

Permet notamment de contrôler ou de mettre en évidence d’éventuelles *variables parasites*.

Interaction entre facteurs

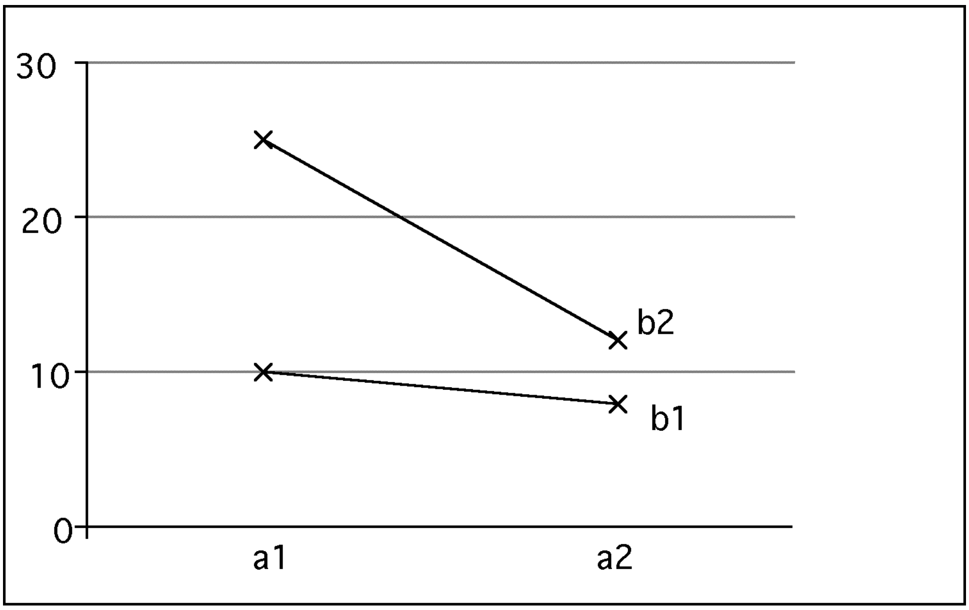
Exemple : Mémorisation d’une liste de mots.
VD : nombre de mots mémorisés

Facteur $\mathcal{A}$: a_1 sujets normaux, a_2 déficit mnésique
Facteur $\mathcal{B}$: b_1 liste de 12 mots, b_2 30 mots

Moyennes observées sur 4 groupes indépendants :

	b_1	b_2
a_1	10	25
a_2	8	12

Graphe d’interaction :



Effet simple d'un facteur $\mathcal{A}$: effet observé de $\mathcal{A}$ lorsque les autres facteurs sont fixés.

Effet principal d'un facteur $\mathcal{A}$: effet observé de $\mathcal{A}$, sans tenir compte des autres conditions.

Définition et écriture d'un plan d'expérience

En général, plusieurs facteurs, avec interaction. Donc : étude simultanée.

Plan factoriel :

Un plan factoriel est un plan dans lequel chaque modalité d'un facteur est combinée avec chaque combinaison de modalités des autres facteurs.

Plan en carré latin

Exemple : 3 facteurs comportant le même nombre de modalités.

On croise les deux premiers facteurs. Les modalités du troisième sont distribuées de façon à réaliser des permutations sur les lignes et les colonnes.

Exemple.

$a_1b_1c_1$	$a_1b_2c_2$	$a_1b_3c_3$
$a_2b_1c_2$	$a_2b_2c_3$	$a_2b_3c_1$
$a_3b_1c_3$	$a_3b_2c_1$	$a_3b_3c_2$

Plans quasi-complets (Rouanet - Lépine 1976)

Croisement : deux ou plusieurs facteurs sont croisés si chaque niveau de l'un des facteurs est combiné avec chaque niveau de chacun des autres facteurs.
Notation : $\mathcal{A}_3 * \mathcal{B}_5$ par exemple.

Emboîtement : Un facteur $\mathcal{A}$ est emboîté dans un facteur $\mathcal{B}$ si chaque niveau de $\mathcal{A}$ est combiné avec un seul niveau de $\mathcal{B}$.
Notation : $\mathcal{A} < \mathcal{B} >$

Emboîtement équilibré : pour chaque niveau du facteur emboîtant, on a le même nombre de niveaux du facteur emboîté.

Les plans *quasi-complets* sont les plans qui peuvent être décrits à l'aide de relations de croisement et d'emboîtement.

Définition :

Un plan est dit *quasi-complet* s'il possède les deux propriétés suivantes :

– Tous les facteurs croisés deux à deux sont croisés ou emboîtés

– Les facteurs croisés deux à deux sont croisés dans leur ensemble.

Exemple : $\mathcal{S}_4 < \mathcal{A}_2 > *\mathcal{B}_2$

Lorsque le facteur *sujet* est croisé avec d'autres facteurs : *plan à mesures répétées*

Modèle de score

Hypothèse : additivité des effets

Pour un plan $\mathcal{S} < \mathcal{A}_a >$:

$$x_{ij} = \mu + \alpha_j + e_{ij}$$

Pour un plan $\mathcal{S} * \mathcal{A}_a$

$$x_{ij} = \mu + \alpha_j + \epsilon_i + e_{ij}$$

Pour un plan $\mathcal{S} < \mathcal{A}_a * \mathcal{B}_b >$:

$$x_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + e_{ijk}$$

$\alpha\beta_{jk}$: effet de l'interaction entre $\mathcal{A}$ et $\mathcal{B}$.

Analyse de variance à plusieurs facteurs

Plan $\mathcal{S}_n * \mathcal{A}_a$

Notations
a : nombre de conditions expérimentales.
n : nombre de sujets.
x_{ij} : valeur de la VD pour le i-ième individu dans la condition expérimentale j.

Hypothèses du test
H_0 : Dans la population parente, les moyennes correspondant aux a conditions expérimentales sont égales.
H_1 : Les moyennes sont différentes.

Présentation des résultats

Source	S. carrés	ddl	C. moyen	F
$\mathcal{A}$	SC_A	$a-1$	CM_A	$\frac{CM_A}{CM_{AS}}$
$\mathcal{S}$	SC_S	$n-1$	CM_S	
Résid.	SC_{AS}	$(n-1)(a-1)$	CM_{AS}	
Total	SC_T	$N-1$		

Les carrés moyens sont obtenus en divisant la somme des carrés de la ligne par le nombre de ddl de la même ligne.
F suit une loi de Fisher-Snedecor à $a-1$ et $(n-1)(a-1)$ degrés de liberté.

Exemple : Effet du bruit sur la discrimination perceptive
Facteur : bruit (3 niveaux)
VD : nombre d'erreurs commises.

Sujets	Absence	Intermittent	Continu
1	117	119	127
2	130	126	131
3	122	118	129
4	123	117	134
5	126	120	137
6	116	120	128

Source	S. carrés	*ddl*	C. moyen	F
Bruit	403.11	2	201.56	19.98 **
Sujets	164.44	5	32.89	
Résid.	100.89	10	10.09	
Total	668.44	17		

Plan $\mathcal{S} < \mathcal{A}_a * \mathcal{B}_b >$

$\mathcal{A}$ et $\mathcal{B}$: facteurs fixes.

Notations
a, b, n, x_{ijk}, N

Interaction entre les facteurs $\mathcal{A}$ et $\mathcal{B}$

Tableau d'analyse de variance

Source	S. carrés	ddl	C. moyen	F
$\mathcal{A}$	SC_A	$a-1$	CM_A	$\frac{CM_A}{CM_{S(AB)}}$
$\mathcal{B}$	SC_B	$b-1$	CM_B	$\frac{CM_B}{CM_{S(AB)}}$
$\mathcal{AB}$	SC_{AB}	$(a-1)(b-1)$	CM_{AB}	$\frac{CM_{AB}}{CM_{S(AB)}}$
Résid.	$SC_{S(AB)}$	$ab(n-1)$	$CM_{S(AB)}$	
Total	SC_T	$N-1$		

Comme précédemment, chaque carré moyen est calculé en divisant la somme des carrés de la ligne par le nombre de ddl correspondant.

Les statistiques F_A, F_B, F_{AB} suivent des lois de Fisher Snedecor, avec des nombres de ddl différents. Le nombre de degrés de liberté du numérateur est respectivement $(a-1)$, $(b-1)$ et $(a-1)(b-1)$. Celui du dénominateur est $ab(n-1)$.

Exemple : facteurs : sexe, statut socio-économique.
VD : mesure du "locus of control"

	statut socio-économique		
	Bas	Moyen	Elevé
Hommes	10	16	18
	12	12	14
	8	19	17
	14	17	13
	10	15	19
	16	11	15
	15	14	22
	13	10	20
Femmes	8	14	12
	10	10	18
	7	13	14
	9	9	21
	12	17	19
	5	15	17
	8	12	13
	7	8	16

Sources de var.	ddl	SC	CM	F
Sexe	1	65.33	65.33	7.73**
Statut soc-éco	2	338.67	169.33	20.03**
$X \times C$	2	18.67	9.33	1.10 NS
Résidu	42	355.0	8.45	
Total	47	777.67		

Conclusion : les deux facteurs (sexe et statut socio-économique) ont des effets significatifs. En revanche, on n'a pas observé d'interaction entre ces facteurs.