II-1. *G* est un (seul) facteur à effets fixes

II-1-1 Généralités

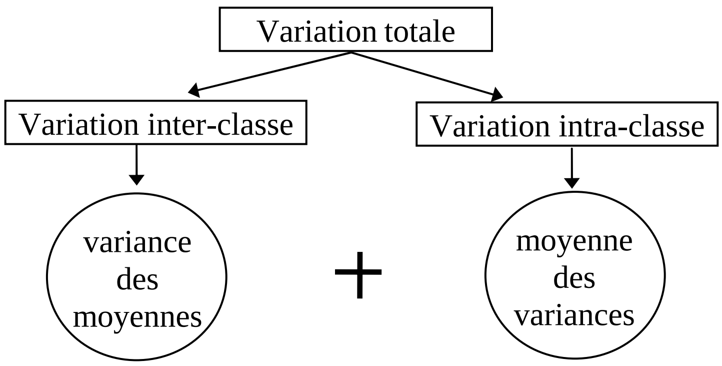
G est un facteur à r modalités (groupes) g_1, g_2, … g_r d'effectifs respectifs n_1, n_2, … n_r .
$N=n_1+n_2+ \ldots +n_r$. Y est la variable dépendante et $y_{s(i)}$ est le score du sujet s dans le groupe g_i (c'est l'observation associée au sujet s dans la modalité g_i).

	Facteur *G* (groupe)						
Sujets	g_1	g_2	…	g_i	…	g_r	
1 2 … s … … …	$y_{1(1)}$ $y_{2(1)}$ … $y_{s(1)}$ … $y_{n_1(1)}$	$y_{1(2)}$ $y_{2(2)}$ … $y_{s(2)}$ … $y_{n_2(2)}$ …	… … … … … … …	$y_{1(i)}$ $y_{2(i)}$ … $y_{s(i)}$ … $y_{n_i(i)}$	… … … … … … …	$y_{1(r)}$ $y_{2(r)}$ … $y_{s(r)}$ … $y_{n_r(r)}$ …	
Effectif	n_1	n_2	…	n_i	…	n_r	$N=\sum_{i=1}^{r} n_i$
Moyenne	$\overline{y}_1$	$\overline{y}_2$	…	$\overline{y}_i$	…	$\overline{y}_r$	$\overline{y}=\frac{1}{N}\sum_{i=1}^{r} n_i\overline{y}_i$
Variance	s'_1	s'_2	…	s'_i	…	s'_r	s'

On veut étudier l'effet du facteur G sur la variable réponse Y.
Pour cela on testera l'hypothèse nulle : **H$_0$** : le facteur n'a pas d'effet sur Y
contre l'hypothèse alternative : **H$_1$** : le facteur a de l'effet sur Y.

II-1-2 Décomposition de la variance (Rappel)

Variance totale = Var Inter + Var Intra



II-1-3 Modèle statistique (Modèle du score, modèle linéaire)

On veut étudier l'effet d'un facteur G à r modalités sur une variable réponse Y. On dispose pour chaque modalité i du facteur G de n_i observations. On note $y_{s(i)}$ l'observation du sujet s dans la modalité i.

$y_{s(i)}$ est la réalisation de la variable aléatoire $Y_{s(i)}$ décrite par le modèle suivant : $Y_{s(i)} = \mu+\alpha_i+\varepsilon_{s(i)}$

où :

- μ s'interprète comme un niveau général.
- α_i mesure l'effet de la $i^{\text{ème}}$ modalité du facteur G.
- les variables aléatoires $\varepsilon_{s(i)}$ sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$
- Pour tout i=1, 2,…, r. Les variances des r groupes sont homogènes (homoscédasticité)
- On rajoute une contrainte d'identifiabilité: $\sum_{i=1}^{r} n_i\alpha_i = 0$ (en moyenne les effets sont nuls)

On peut écrire l'hypothèse nulle et l'hypothèse alternative sous la forme suivante :

hypothèse nulle : **H₀** : $\alpha_i=0$ pour tout i = 1, 2,…, r

hypothèse alternative : **H₁** : l'un au moins des α_i est non nul

On montre alors les résultats suivants :

Théorème :

- E(**SCR**) = $(N - r)\sigma^2$
- E(**SCG**) = $(r-1)\sigma^2 + \sum_i n_i\alpha_i^2$

Remarque : Il est clair que les valeurs des variations (totale, inter et intra) dépendent des effectifs des modalités du facteur G. Aussi pour apprécier plus justement les grandeurs relatives à ces variations (en particulier pour comparer la variation due au facteur à la variation résiduelle), on calcule les carrés moyens des écarts, en divisant chaque variation par son nombre de degrés de liberté (ddl) : $\mathbf{MC} = \dfrac{\mathbf{SC}}{\text{ddl}}$

- carrés moyens (total) : **MCT**

  La variation totale concerne N scores (observations) $y_{s(i)}$ liés par une relation $\bar{y} = \dfrac{1}{N}\sum_{i,j} y_{s(i)}$ donc présente N - 1 ddl et on a : $\mathbf{MCT} = \dfrac{\mathbf{SCT}}{N-1}$

- carrés moyens inter : **MCG**

  La variation inter(-groupes) concerne r groupes liés par une relation $\bar{y} = \dfrac{1}{N}\sum_i n_i\bar{y}_i$ donc présente r - 1 ddl et on a : $\mathbf{MCG} = \dfrac{\mathbf{SCG}}{r-1}$

- carrés moyens intra (résiduels) : **MCR**

  La variation intra (résiduelle) concerne N scores (observations) liés par r relations $\bar{y}_i = \dfrac{1}{n_i}\sum_j y_{ji}$ donc présente N - r ddl et on a : $\mathbf{MCR} = \dfrac{\mathbf{SCR}}{N-r}$

Remarque : Tout comme les variations, les degrés de liberté sont additifs

ddl total = ddl inter + ddl intra $N - 1 = r - 1 + N - r$

Mais les carrés moyens ne sont pas additifs. En général **MCT** $\neq$ **MCG** + **MCR**.

<u>Le théorème précédent</u> peut se réécrire : E(**MCR**) = σ^2 E(**MCG**) = $\sigma^2 + \sum_{i=1}^{r} \frac{n_i \alpha_i^2}{r-1}$

Par conséquent : MCR est une estimation sans biais de la variance σ^2 et **MCG** est une estimation de σ^2 augmentée d'un terme positif fonction des effets de groupe ($\sum_{i=1}^{r} \frac{n_i \alpha_i^2}{r-1}$).

Sous l'hypothèse nulle ($\alpha_i = 0$ pour tout $i = 1,2,\ldots,r$) **MCG** est aussi une estimation de σ^2.

Théorème :
Dans le cadre du modèle statistique, la statistique $\frac{\mathbf{SCR}}{\sigma^2}$ suit une loi du χ^2 à $N - r$ ddl et sous l'hypothèse nulle la statistique $\frac{\mathbf{SCG}}{\sigma^2}$ suit une loi du χ^2 à $r - 1$ ddl et ces deux statistiques sont indépendantes.

Corollaire :
Sous l'hypothèse nulle la statistique $\frac{\mathbf{SCG}/(r-1)}{\mathbf{SCR}/(N-r)} = \frac{\mathbf{MCG}}{\mathbf{MCR}} = \mathbf{F_G}$ suit une <u>loi de Fischer</u> à $r - 1$ et $N - r$ ddl, notée $\mathcal{F}(r - 1, N - r)$.

Conclusion :
Le test d'hypothèse est alors défini au seuil de signification α par la règle de décision suivante :
Si $F_{obs} \geq \lambda_\alpha$ alors on rejette l'hypothèse nulle. λ_α étant donné par l'équation : $\alpha = P(F \geq \lambda_\alpha)$.

II-1-4 Tableau d'analyse de la variance

Les résultats sont souvent présentés sous forme d'un tableau :

Source de variation	Somme des carrés des écarts observés SC_{obs}	ddl	Carrés moyens observés MC_{obs}	F
G, Facteur, Inter-groupes, ...	**SCG**	r - 1	**MCG**	$\mathbf{F_{obs}} = \frac{\mathbf{MCG}}{\mathbf{MCR}}$
S<G>, Résiduelle, Intra	**SCR**	N - r	**MCR**	F_{lu} = α =
Totale	**SCT**	N - 1		

II-1-5 Exemple :

Pour étudier l'influence du facteur « intensité du bruit environnant » sur la capacité d'un sujet à résoudre un problème, l'expérimentateur construit l'expérience suivante : 24 écoliers sont répartis de façon aléatoire dans quatre pièces. Des bruits de la rue ont été enregistrés et sont diffusés dans chaque pièce avec un niveau sonore particulier. Les enfants doivent résoudre une série de problèmes. La variable réponse est la note finale obtenue à la série d'épreuves.

	Niveau sonore				
	1	2	3	4	
	62 60 63 59	56 62 60 61 63 64 63 59	63 67 71 64 65 66	68 66 71 67 68 68	
n_i	4	8	6	6	$\boldsymbol{N} = 24$
$\overline{y}_i$	61	61	66	68	$\overline{y} = 64$
Variance	10/4	48/8	40/6	14/6	SCIntra= **SCR** = 112
$n_i\left(\overline{y}_i - \overline{y}\right)^2$	36	72	24	96	SCInter= $\sum_{i=1}^{r} n_i\left(\overline{y}_i - \overline{y}\right)^2$ =228

SCT = $N\times$variance totale = 340 $\quad$ $\textbf{MCG} = \frac{\textbf{SCG}}{r-1} = \frac{228}{3} = 76$ $\quad$ $\textbf{MCR} = \frac{\textbf{SCR}}{N-r} = \frac{112}{20} = 5{,}6.$

On peut présenter les résultats dans le tableau d'analyse de variance :

Sources de variations	SC_{obs}	ddl	MC_{obs}	Statistique de test = Q
Facteur = Inter = G	*228*	r-1 = 3	76	$\mathbf{F}_{obs} = \frac{\mathbf{MCG}_{obs}}{\mathbf{MCR}_{obs}} = \frac{76}{5{,}6} = 13{,}5714^{**}$
$S(G)$ = Intra	*112*	N-r = 20	5,6	-
Total	*340*	N-1 = 23	-	-

En prenant un niveau de signification α=0,05=5%, on peut lire la valeur de λ_α dans la table de Fischer à r-1 = 3 et N-r = 20 degrés de liberté : $\lambda_\alpha = \mathcal{F}_{(3\,;\,20\,;\,0{,}95)} = 3{,}10$. Donc, si $\mathbf{F}_{obs}$, la valeur observée de la statistique de test, est supérieure ou égale à 3,10 on rejette $\mathbf{H_0}$ et dans le cas contraire, on conserve l'hypothèse $\mathbf{H_0}$.

Conclusion *:* Comme $\mathbf{F}_{obs}$=13,5714 > λ_α=3,10, on rejette $\mathbf{H}_0$ pour accepter $\mathbf{H}_1$; C'est-à-dire qu'il y a effectivement un effet du bruit environnant sur la capacité de résolution des problèmes. On rejette $\mathbf{H_0}$ même avec un α=0,01=1%.

II-2. *G* est un (seul) facteur à effets aléatoires

II-2-1 Modèle statistique

On choisit un échantillon de r modalités $g_1, g_2, \dots g_r$ du facteur G. On dispose pour chaque modalité échantillonnée de n_i observations de la variable dépendante ($N=n_1+n_2+ \dots +n_r$). On note $y_{s(i)}$ l'observation du sujet s dans le groupe g_i. Chaque observation $y_{s(i)}$ est la réalisation d'une variable aléatoire $Y_{s(i)}$ décrite par le modèle : $Y_{s(i)} = \mu + \Gamma_i + \varepsilon_{s(i)}$

où :

μ est une constante mesurant le niveau général de la réponse

Γ_i est une variable aléatoire qui mesure l'effet aléatoire de G.

$\varepsilon_{s(i)}$ est une variable aléatoire représentant le résidu.

On suppose réalisées les trois hypothèses suivantes :

- les $\varepsilon_{s(i)}$ sont des variables aléatoires indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$
- les Γ_i (i=1, 2, …, r) sont des variables aléatoires indépendantes de même loi $\mathcal{N}(0, \sigma_G^2)$
- les Γ_i sont indépendantes des $\varepsilon_{s(i)}$.

Le test d'hypothèse s'écrira alors :

H_0 : G n'a pas d'effet $\Leftrightarrow \sigma_G^2 = 0$

H_1 : G a un effet $\Leftrightarrow \sigma_G^2 > 0$

On montre alors les résultats suivants :

Théorème :

- E(**MCR**) = σ^2 ou encore E(**SCR**) = $(N - r)\sigma^2$
- E(**MCG**) = $\sigma^2 + K\sigma_G^2$ où K est une constante.

II-2-2 Test statistique

On se ramène donc au cas précédent et on utilisera la même statistique $\mathbf{F} = \dfrac{\mathbf{MCG}}{\mathbf{MCR}}$ pour réaliser le test.

II-2-3 Exemple :

On veut vérifier que l'intensité de traitement perceptif d'un visage dépend du visage examiné (certains visages retiennent plus l'attention que d'autres). Pour mettre à l'épreuve cette hypothèse de recherche, on construit l'expérience suivante :

40 sujets sont choisis au hasard et répartis de façon aléatoire dans 5 groupes de 8 sujets chacun. Chaque groupe examine un visage choisi par l'expérimentateur au hasard dans l'ensemble de visages disponibles.

L'expérimentateur mesure l'intensité du traitement perceptif en observant la dilatation de la pupille lors de l'examen du visage. Il obtient les résultats suivants :

Sujets	Groupe des visages					
	g_1	g_2	g_3	g_4	g_5	
1	58	60	63	64	57	
2	51	61	55	64	59	
3	57	66	57	65	65	
4	59	65	60	61	63	
5	56	59	61	66	62	
6	54	59	62	59	64	
7	53	64	58	67	60	
8	52	63	56	60	63	
n_j	8	8	8	8	8	N=40
$\bar{y}_i$	55	62,125	59	63,25	61,625	$\bar{y}$ = 60,2
s'^2_i	7,5	6,6094	7,5	7,4375	6,4844	s'^2=15,81

SCT = $N\,s'^{\mathbf{2}}$ = 40 × 15,81 = 632,4

SCG = $N\ s'^2_{\bar{y}_i} = N\ \sum_{i=1}^{5}\left(\bar{y}_i - \bar{y}\right)^2$ = 40 × 8,7038 = 348,15

SCR = $N\ \sum_{i=1}^{5} s_i'^2$ = 8 × [7,5+6,6094+7,5+7,4375+6,4844] = 8 × 35,5313 = 284,25

On vérifie que : **SCR = SCT - SCG** = 632,4 - 348,15 = 284,25**.**

On peut présenter les résultats dans le tableau d'analyse de variance :

Sources de variations	SC_{obs}	ddl	MC_{obs}	Statistique de test = Q
Facteur = Inter = G	*348,15*	r-1 = 4	87,0375	$\mathbf{F}_{obs}=\dfrac{\mathbf{MCG}_{obs}}{\mathbf{MCR}_{obs}}=\dfrac{87{,}0375}{8{,}1214}=10{,}717^{**}$
$S(G)$ = Intra	*284,25*	N-r = 35	8,1214	-
Total	*632,4*	N-1 = 39	-	-

En prenant un niveau de signification α=0,05=5%, on peut lire la valeur de λ_α dans la table de Fischer à r-1 = 4 et N-r = 35 degrés de liberté : $\lambda_\alpha = \mathcal{F}_{(4\,;\,35\,;\,0{,}95)} = 2{,}64$

Conclusion : Comme $\mathbf{F}_{obs}=10{,}717 > \lambda_\alpha=2{,}64$, on rejette $\mathbf{H}_0$ pour accepter $\mathbf{H}_1$; C'est-à-dire qu'il y a effectivement un effet du facteur visage. ($\sigma_G^2 > 0$)

Remarque : En prenant un niveau de signification α=0,01=1%, on peut lire la valeur de λ_α dans la table de Fischer à r-1 = 4 et N-r = 35 degrés de liberté : $\lambda_\alpha = \mathcal{F}_{(4\,;\,35\,;\,0{,}99)} = 3{,}91$ et on tire la même conclusion.