Analyse de Variance à 1 Facteur

Supposons que l'on étudie les scores obtenus à une épreuve de math (variable dépendante) par 4 classes de seconde de 4 lycées différents (variable indépendante).

Groupes	A	B	C	D	
Scores	* * * *	* * * * *	* * *	* * * *	
Effectif	n_A	n_B	n_C	n_D	$N = n_A + \dots + n_D$
Total : Σx_i	T_A	T_B	T_C	T_D	$T_G = T_A + \dots + T_D$
Moyenne	$\bar{x}_A$	$\bar{x}_B$	$\bar{x}_C$	$\bar{x}_D$	$\bar{x}_T = T_G / N$

Table de calcul de l'analyse de variance

Question: Les scores sont-ils différents entre les Lycées?

Les deux conditions suivantes sont supposées vérifiées :

La variable (note en math) se distribue normalement dans les ensembles parents des quatre classes.

Les ensembles parents des quatre classes ont les mêmes variances. Ces classes ne peuvent donc différer entre elles que par leurs moyennes.

Ho : Les échantillons de scores de moyennes ($\bar{x}_A$, ..., $\bar{x}_D$) sont issus d'une même population parente de moyenne μ_T

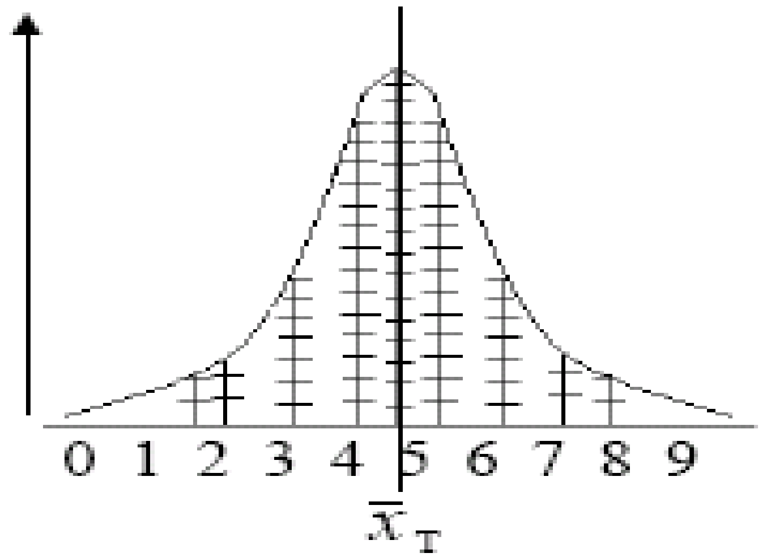
$$\mu_A = \mu_B = \mu_C = \mu_D = \mu_T$$

H1 : Les quatre groupes sont extraits de populations parentes de moyennes différentes. Au moins l'une des moyennes d'échantillons est différente de μ_T

Ho teste donc l'hypothèse de l'homogénéité des k moyennes. On dit aussi que le facteur (à partir duquel on a construit les k groupes) n'a pas d'influence sur la variable X (note en math).

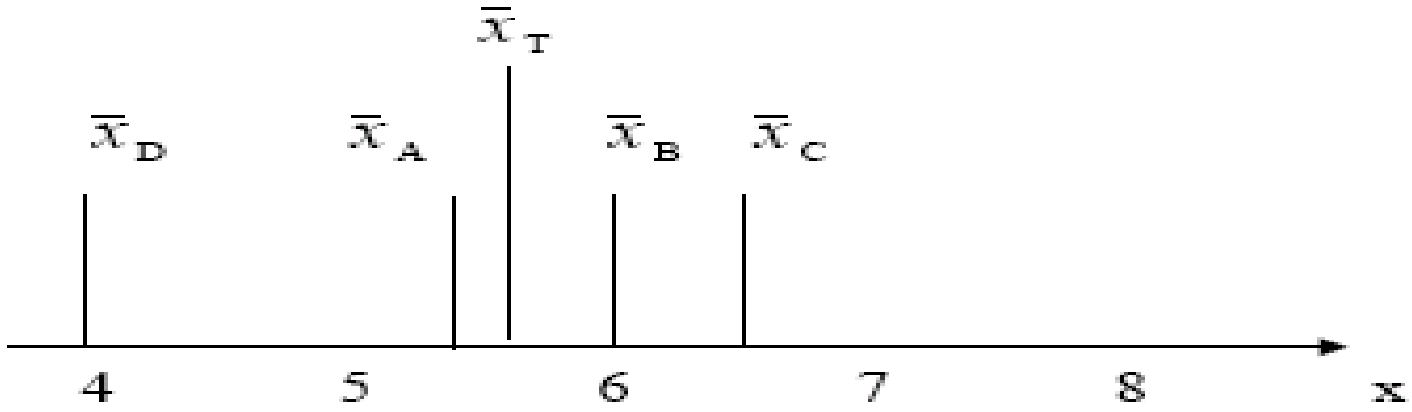
1) La variation totale de la variable X est mesurée par la dispersion de toutes les valeurs (N valeurs) autour de la moyenne générale $\overline{x}_T$ de la distribution de l'ensemble des notes des quatre classes.

Effectifs



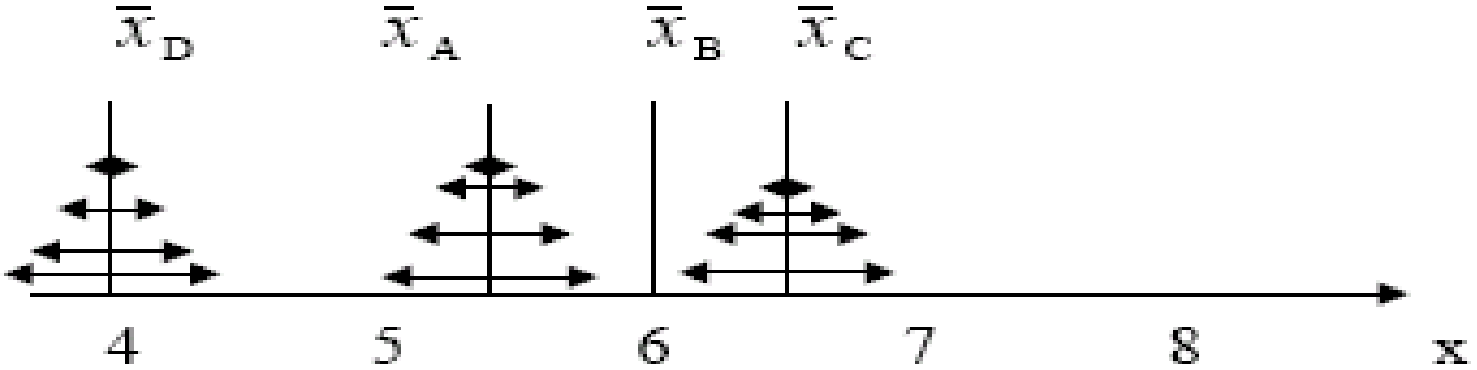
2) La variation entre les classes (variation inter-groupe)

Chaque classe est caractérisée par sa moyenne ($\bar{x}_A$, ..., $\bar{x}_D$) qui s'écarte plus ou moins de la moyenne générale ($\bar{x}_T$).



3) La variation à l'intérieur de chaque classe (intra-groupe, résiduelle)

Cette variation résiduelle est mesurée par la Somme des Carrés des écarts de chacune des valeurs de la moyenne de son groupe.



$$SC_T = SCg + SCr$$

Variation totale de la variable x

Variation résiduelle (intra groupe)

Variation inter-groupe (Effet du facteur étudié : variable indépendante)

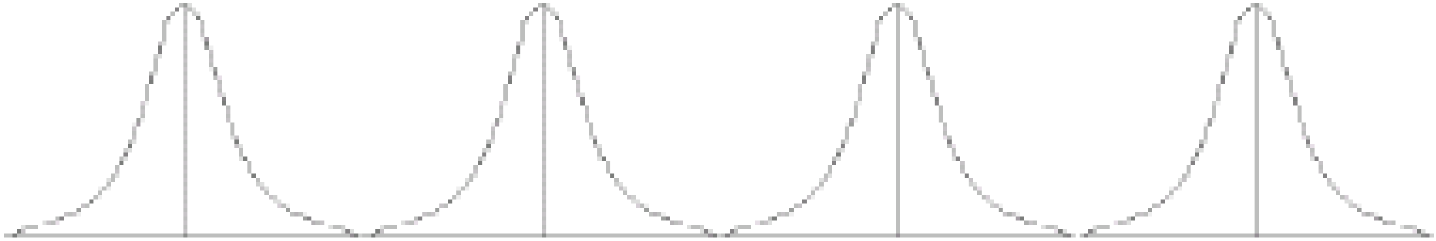
Les valeurs de ces trois variations dépendent des effectifs des groupes. Aussi, pour apprécier les grandeurs relatives de ces variations, on calcule pour chacune les Carrés Moyens : on divise chaque variation (somme des carrés des écarts) par son nombre de degré de liberté.

Snedecor s'est intéressé au rapport de grandeur entre la valeur du carré moyen inter-groupes (dû au facteur) et la valeur du carré moyen résiduel.
Ce rapport, c'est la statistique de décision utilisée en analyse de variance, à savoir le F.

$$\mathbf{F} = \frac{\textbf{Variance inter}}{\textbf{Variance intra}} = \frac{\textbf{CMg}}{\textbf{CMr}}$$

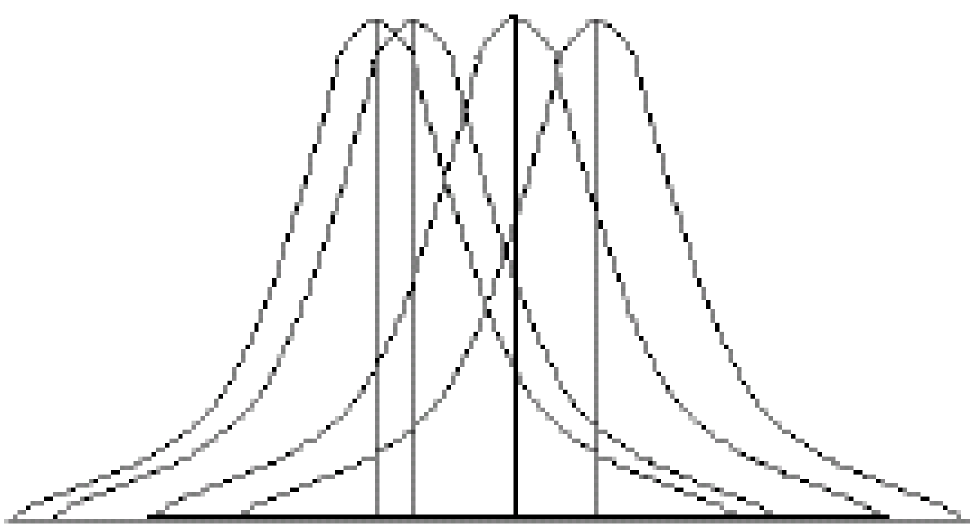
1) Variance Inter-groupes élevée et Variance Intra-groupe faible

$$\Rightarrow \quad \frac{CMg}{CMr} \nearrow \qquad F \nearrow$$



2) Variance Inter-groupes faible et Variance Intra- groupe élevée

$$\Rightarrow \quad \frac{CMg}{CMr} \searrow \qquad F \searrow$$



Source de variation	Somme des carrés des écarts	Nombre de ddl	Carrés Moyens (variances)	F
entre les groupes (Inter)				
à l'intérieur des groupes (intra)				
Total				

Avec $F =$