

Rapport de Stage

Thème du Stage

Prénom Nom

votre.email@example.com

13 juin 2024

Table des matières

0.1	Introduction	2
0.2	Présentation théorique	2
0.2.1	Analyse de variance à 2 facteurs	2
1	Conclusion	6

0.1 Introduction

L'éducation physique et sportive (EPS) joue un rôle crucial dans le développement physique et mental des élèves, en leur offrant des occasions de pratiquer une activité physique régulière et structurée. Cependant, il existe des disparités notables dans l'engagement physique des élèves en fonction de divers facteurs.

0.2 Présentation théorique

0.2.1 Analyse de variance à 2 facteurs

L'analyse de variance (ou ANOVA) est une méthode permettant de modéliser la relation entre une variable quantitative et une ou plusieurs variables qualitatives. Quand il y a une seule variable explicative, on parle d'analyse de variance à un facteur. Cette méthode permet de comparer k ($k \geq 2$) moyennes et peut donc être vue comme une extension du test de comparaison de moyenne.

Ici nous allons nous intéresser au cas de deux variables explicatives qualitatives, donc à l'analyse de variance à deux facteurs, qu'on peut toutefois généraliser à un nombre quelconque de variables explicatives.

Modèle d'analyse de variance (ANOVA à deux facteurs)

Notons A et B deux variables explicatives qualitatives, et Y une variable explicative quantitative. Soit I le nombre de modalités de A et J celui de la variable B. Le modèle s'écrit classiquement :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K, \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2), \quad (1)$$

avec un effet moyen général μ , α_i un effet de la modalité i du facteur A, β_j un effet de la modalité j du facteur B, un terme d'interaction γ_{ij} et un résidu ϵ_{ijk} avec k l'indice de répétition pour le couple (i, j) .

Le modèle n'est pas identifiable car pour un $(1+I+J+IJ)$ -uplet $(\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, \gamma_{11}, \dots, \gamma_{IJ}, \dots)$ et $a \in \mathbf{R}$, le $(1+I+J+IJ)$ -uplet $(\mu - a, \alpha_1 + \frac{a}{3}, \dots, \alpha_I + \frac{a}{3}, \beta_1 + \frac{a}{3}, \dots, \beta_J + \frac{a}{3}, \gamma_{11} + \frac{a}{3}, \dots, \gamma_{IJ} + \frac{a}{3}, \dots)^T$ correspond au même modèle. On a donc besoin de contrainte sur les paramètres. Les contraintes classiques sont :

— Contrainte de type analyse par cellule

$$\mu = 0, \quad \forall i \quad \alpha_i = 0, \quad \forall j \quad \beta_j = 0. \quad (2)$$

— Contrainte de type cellule de référence

$$\alpha_1 = 0, \quad \beta_1 = 0, \quad \forall i \quad \gamma_{i1} = 0, \quad \forall j \quad \gamma_{1j} = 0 \quad (3)$$

— Contrainte de type somme

$$\sum_i \alpha_i = 0, \quad \sum_j \beta_j = 0, \quad \forall i \quad \sum_j \gamma_{ij} = 0, \quad \forall j \quad \sum_i \gamma_{ij} = 0 \quad (4)$$

Conditions d'application

Les conditions d'application de l'ANOVA sur nos données sont :

- Indépendance de nos observations.
- Égalité de la variance des résidus dans les différents groupes (homoscédasticité).
- Normalité des résidus du modèle.

Test de l'interaction et de significativité des effets de facteur

Pour tester l'effet d'un facteur, il est naturel de tester la nullité des paramètres associés à chaque niveau de ce facteur. Cette approche est insuffisante pour définir correctement l'hypothèse nulle testée, il faut préciser les hypothèses portant sur les autres facteurs et interactions. Si le plan est équilibré, quel que soit les hypothèses portant sur les autres variables, le test de significativité de facteur reste identique pour le même facteur mais dans le cas de plan dés-équilibré ce test est parfois différents en fonction des hypothèses portant sur les autres facteurs et interactions. Donc on utilise la notion de réduction, définie ci-dessous, qui permet de bien spécifier le rôle des autres facteurs et interactions.

Définition : Soit un modèle contenant les effets (a_1, \dots, a_l) des niveaux de facteurs $(Var_1, Var_2, \dots, Var_l)$. On appelle réduction associée à l'introduction de a_{q_1}, \dots, a_{q_d} dans un modèle contenant les effets a_{i_1}, \dots, a_{i_m} , notée $R(a_{q_1}, \dots, a_{q_d} | \mu, a_{i_1}, \dots, a_{i_m})$ la norme suivante :

$$R(a_{q_1}, \dots, a_{q_d} | \mu, a_{i_1}, \dots, a_{i_m}) = SCE_{i_1, i_2, \dots, i_m, q_1, q_2, \dots, q_d} - SCE_{i_1, i_2, \dots, i_m}, \quad (5)$$

avec SCE_{i_1, i_2} la somme des carrés expliquée par le modèle ne contenant que les niveaux de facteurs Var_{i_1}, Var_{i_2} .

Soit $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_I)$, $\beta = (\beta_1, \beta_2, \dots, \beta_J)$ et $\gamma = (\gamma_{ij})$ avec $i = 1, \dots, I$, $j = 1, \dots, J$, les effets respectifs de niveaux de facteurs des modalités A, B et de l'interaction.

La variabilité de Y se décompose en la somme de deux termes :

$$SCT = SCE + SCR \quad (6)$$

On a donc

$$R(\alpha | \mu, \beta, \gamma) = SCE_1 - SCE_2, \quad (7)$$

où SCE_1 et SCE_2 sont respectivement la somme des carrés expliquée des modèles

$$(M_1) : Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

$$(M_2) : Y_{ijk} = \mu + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

pour tout $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$.

Un test de l'effet d'un facteur est associé à une réduction donnée. Considérons la réduction $R(\alpha | \mu, \beta, \gamma)$, les hypothèses associée s'écrit :

- H_0 = Il n'y a pas de différence significative entre le modèle ne contenant pas l'effet du facteur A et le modèle contenant les effets du facteur A, B et l'interaction.
- H_1 = Il y a une différence significative entre le modèle ne contenant pas l'effet du facteur A et le modèle contenant les effets du facteur A, B et l'interaction.

La statistique de test est donnée par :

$$F = \frac{\frac{R(\alpha|\mu, \beta, \gamma)}{I-1}}{\frac{SCR}{n-IJ}}, \quad (8)$$

où

- $I - 1$ est le nombre de contraintes dans l'hypothèse nulle.
- IJ est le nombre de paramètres de notre modèle.

Sous H_0 , $F \sim \mathcal{F}(I - 1, n - IJ)$.

Quelques réductions classiques :

Parmi les plus classiques, on trouve les réductions de type I, type II et type III.

a) Somme des carrés de type I

La réduction de type I provient de la décomposition de la somme des carrés expliquée par le modèle en réductions successives. Dans le modèle M_1 ci-dessus,

$$SCE_1 = R(\alpha, \beta, \gamma|\mu) = R(\alpha|\mu) + R(\beta|\alpha, \mu) + R(\gamma|\alpha, \beta, \mu) \quad (9)$$

Le tableau d'analyse de variance correspondant au test est le suivant :

Effet	Réduction type I	DDL	F	Question
α	$R(\alpha \mu)$	$I - 1$	$\frac{R(\alpha \mu)}{I-1} \frac{SCR}{n-IJ}$	Est-il pertinent d'ajouter l'effet du facteur A à un modèle ne contenant que la constante ?
β	$R(\beta \mu, \alpha)$	$J - 1$	$\frac{R(\beta \mu, \alpha)}{J-1} \frac{SCR}{n-IJ}$	Est-il pertinent d'ajouter l'effet du facteur B à un modèle contenant la constante et l'effet du facteur A ?
γ	$R(\gamma \mu, \alpha, \beta)$	$(I - 1) \times (J - 1)$	$\frac{R(\gamma \mu, \alpha, \beta)}{(I-1) \times (J-1)} \frac{SCR}{n-IJ}$	Est-il pertinent d'ajouter l'effet d'interaction entre les deux facteurs à un modèle contenant la constante et les effets des deux facteurs ?

TABLE 1 – Table d'analyse de la variance des réductions de type I du modèle M_1 .

b) Somme des carrés de type II

Dans la réduction de type II l'ordre d'introduction des facteurs dans le modèle leur confère un rôle différent. L'idée des types II consiste à considérer la réduction portée par un facteur conditionnellement aux autres. Donc l'ordre n'est plus important.

Le tableau d'analyse de variance correspondant au test est le suivant :

Effet	Réduction type II	DDL	F	Question
α	$R(\alpha \beta, \mu)$	$I - 1$	$\frac{R(\alpha \beta, \mu)}{I-1} \frac{SCR}{(n-IJ)}$	Est-il pertinent d'ajouter l'effet du facteur A à un modèle contenant la constante et l'effet du facteur B ?
β	$R(\beta \mu, \alpha)$	$J - 1$	$\frac{R(\beta \mu, \alpha)}{J-1} \frac{SCR}{(n-IJ)}$	Est-il pertinent d'ajouter l'effet du facteur B à un modèle contenant la constante et l'effet du facteur A ?
γ	$R(\gamma \mu, \alpha, \beta)$	$(I - 1) \times (J - 1)$	$\frac{R(\gamma \mu, \alpha, \beta)}{(I-1) \times (J-1)} \frac{SCR}{(n-IJ)}$	Est-il pertinent d'ajouter l'effet de l'interaction entre les deux facteurs à un modèle contenant la constante et les effets des deux facteurs ?

TABLE 2 – Table d'analyse de la variance des réductions de type II du modèle M_1 .

c) Somme des carrés de type III

Si le test portant sur l'interaction est significatif, on utilise donc la réduction de type III, qui consiste à considérer la réduction portée par un facteur conditionnellement aux autres et à l'interaction.

Le tableau d'analyse de variance correspondant au test est le suivant :

Effet	Réduction type III	DDL	F	Question
α	$R(\alpha \beta, \mu, \gamma)$	$I - 1$	$\frac{R(\alpha \beta, \mu, \gamma)}{SCR/(n-IJ)}$	Est-il pertinent d'ajouter l'effet du facteur A à un modèle contenant la constante, l'effet du facteur B et l'interaction ?
β	$R(\beta \mu, \alpha, \gamma)$	$J - 1$	$\frac{R(\beta \mu, \alpha, \gamma)}{SCR/(n-IJ)}$	Est-il pertinent d'ajouter l'effet du facteur B à un modèle contenant la constante, l'effet du facteur A et l'interaction ?
γ	$R(\gamma \mu, \alpha, \beta)$	$(I - 1) \times (J - 1)$	$\frac{R(\gamma \mu, \alpha, \beta)}{SCR/(n-IJ)}$	Est-il pertinent d'ajouter l'effet de l'interaction entre les deux facteurs à un modèle contenant la constante et les effets des deux facteurs ?

TABLE 3 – Table d'analyse de la variance des réductions de type III du modèle M1.

Test de comparaison multiple

Si aucun des facteurs ni l'interaction n'est significatif, l'étude sera conclue. Dans le cas où au moins l'un des facteurs A ou B ou l'interaction est significatif il pourrait être intéressant de comparer les moyennes entre les différents groupes afin de voir ce qui rend le facteur significatif.

Dans le cas où par exemple l'interaction est significatif les test auront donc les hypothèses :

- $H_0 : \theta_{i_1 j_1} = \theta_{i_2 j_2}$
- $H_1 : \theta_{i_1 j_1} \neq \theta_{i_2 j_2}$

où $\theta_{i_1 j_1}$ et $\theta_{i_2 j_2}$ sont les moyennes dans chaque groupes différents et $i_1, i_2 \in 1, \dots, I$; $j_1, j_2 \in 1, \dots, J$.

On veut comparer tous les groupes deux à deux. En comparant m groupes, on effectue $\frac{m(m-1)}{2}$ comparaison à un niveau α . Ce qui nous expose plus au risque de commettre une erreur de type I (Accepter à tort H_1 à un niveau α). Donc pour y remédier nous allons utiliser la méthode de Tukey pour un plan équilibré et Tukey-Kramer pour un plan complet(parfois déséquilibré).

Méthode de Tuckey(-Kramer) :

Pour les tests de comparaison des m moyennes deux à deux l'intervalle de confiance de Tukey(-Kramer) de niveau $1 - \alpha$ pour $\theta_{i_1 j_1} - \theta_{i_2 j_2}$ est

$$\left[\bar{Y}_{i_1 j_1} - \bar{Y}_{i_2 j_2} \pm r_{(1-\alpha)}^{(m, n-m)} \sqrt{\frac{\hat{\sigma}^2}{2} \times \left(\frac{1}{n_{i_1 j_1}} + \frac{1}{n_{i_2 j_2}} \right)} \right] \quad (10)$$

où $r_{(1-\alpha)}^{(m, n-m)}$ est le quantile d'ordre $1-\alpha$ d'une loi spécifique à ce problème, la Studentized range distribution.

Bien sûr il y a d'autre méthode qui ne seront pas abordé dans mon travail tels que : **Méthode de Scheffé, Méthode de Holm-Bonferroni, Correction de Bonferroni,...**

Taille d'effet

Conclusion