

Rapport de Stage

Thème du Stage

Prénom Nom

votre.email@example.com

9 juillet 2024

Table des matières

1	Introduction	3
2	Présentation théorique	3
2.1	Analyse de variance à 2 facteurs	3
2.1.1	Modèle d'analyse de variance (ANOVA à deux facteurs)	3
2.1.2	Conditions d'application	4
2.1.3	Test de l'interaction et de significativité des effets de facteur	4
2.1.4	Test de comparaison multiple	6
2.1.5	Taille d'effet	7
3	Contexte de l'étude et Problématique	8
4	Partie Résultats	9
4.1	Analyse descriptive	9
4.2	Modèles d'ANOVA	10
4.2.1	Examen de l'écart de MVPA entre Filles et Garçons selon le CA	10
4.2.2	Examen de l'écart de MVPA entre Filles et Garçons	16
4.2.3	Examen de l'écart de MVPA entre Filles et Garçons selon la catégorie IPS	16
4.2.4	Examen de l'écart de MVPA entre Filles et Garçons selon le milieu géographique	19
5	Discusion sur les limites des résultats	24
6	Conclusion	24

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.

Présentation théorique

2.1 Analyse de variance à 2 facteurs

L'analyse de variance (ANOVA) est une méthode permettant de modéliser la relation entre une variable quantitative et une ou plusieurs variables qualitatives. Lorsqu'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 moyennes.

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

2.1.1 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 $(\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, \gamma_{11}, \dots, \gamma_{ij}, \dots, \gamma_{IJ})$ et $a \in \mathbf{R}$, le $(1+I+J+IJ)$ -uplet $(\mu - a, \alpha_1 + \frac{a}{3}, \dots, \alpha_I + \frac{a}{3}, \beta_1 + \frac{a}{3}, \dots, \beta_J + \frac{a}{3}, \gamma_{11} + \frac{a}{3}, \dots, \gamma_{ij} + \frac{a}{3}, \dots, \gamma_{IJ} + \frac{a}{3})^T$ correspond au même modèle. On a donc besoin de contraintes 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

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

2.1.2 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.

2.1.3 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 déséquilibré ce test est parfois différent en fonction des hypothèses portant sur les autres facteurs et interactions. On utilise donc 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 α, β et γ les termes respectifs des deux facteurs principaux A, B et de l'interaction A*B.

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 associées s'écrivent :

- 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. Elle consiste à ajouter au modèle un des termes à la fois. 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	$R(\alpha \mu)$	$I - 1$	$\frac{R(\alpha \mu)}{\frac{I-1}{SCR}}$	Est-il pertinent d'ajouter l'effet du facteur A à un modèle ne contenant que la constante ?
β	$R(\beta \mu, \alpha)$	$J - 1$	$\frac{R(\beta \mu, \alpha)}{\frac{J-1}{SCR}}$	Est-il pertinent d'ajouter l'effet du facteur B à un modèle contenant la constante et l'effet du facteur A ?
γ	$R(\gamma \mu, \alpha, \beta)$	$(I - 1) \times (J - 1)$	$\frac{R(\gamma \mu, \alpha, \beta)}{\frac{(I-1) \times (J-1)}{SCR}}$	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 I, l'ordre d'introduction des facteurs dans le modèle leur confère un rôle différent ce qui ne représente aucun intérêt lorsque l'ordre n'est pas important pour notre sujet d'étude. L'idée des types II dans le cadre d'une ANOVA à deux facteurs, consiste à considérer la réduction portée par un facteur ou interaction conditionnellement aux autres facteurs (principaux). Donc l'ordre n'est plus important mais aussi il prend en compte le principe qu'on ne peut pas supprimer l'effet d'un facteur principal du modèle sachant qu'il y a effet de l'interaction. Le tableau d'analyse de variance correspondant au test est le suivant :

Effet	Réduction type II	DDL	F	Question
α	$R(\alpha \beta, \mu)$	$I - 1$	$\frac{R(\alpha \beta, \mu)}{\frac{I-1}{\frac{SCR}{(n-IJ)}}}$	Est-il pertinent d'ajouter l'effet du facteur A à un modèle contenant la constante et l'effet du facteur B ?
β	$R(\beta \mu, \alpha)$	$J - 1$	$\frac{R(\beta \mu, \alpha)}{\frac{J-1}{\frac{SCR}{(n-IJ)}}}$	Est-il pertinent d'ajouter l'effet du facteur B à un modèle contenant la constante et l'effet du facteur A ?
γ	$R(\gamma \mu, \alpha, \beta)$	$(I - 1) \times (J - 1)$	$\frac{R(\gamma \mu, \alpha, \beta)}{\frac{(I-1) \times (J-1)}{\frac{SCR}{(n-IJ)}}}$	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 M1.

c) Somme des carrés de type III

La réduction de type III consiste à considérer la réduction portée par un facteur ou interaction conditionnellement aux autres facteurs ou à l'interaction. Le tableau d'analyse de variance correspondant au test est le suivant :

Effet	Réduction type III	DDL	F	Question
α	$R(\alpha \beta, \mu, \gamma)$	$I - 1$	$\frac{R(\alpha \beta, \mu, \gamma)}{\frac{I-1}{\frac{SCR}{n-IJ}}}$	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	$R(\beta \mu, \alpha, \gamma)$	$J - 1$	$\frac{R(\beta \mu, \alpha, \gamma)}{\frac{J-1}{\frac{SCR}{(n-IJ)}}}$	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	$R(\gamma \mu, \alpha, \beta)$	$(I - 1) \times (J - 1)$	$\frac{R(\gamma \mu, \alpha, \beta)}{\frac{(I-1) \times (J-1)}{\frac{SCR}{(n-IJ)}}}$	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 3 – Table d'analyse de la variance des réductions de type III du modèle M1.

NB : L'ANOVA de type II sera préféré à celui de type III dans toute la suite dans le cas d'un plan déséquilibré.

2.1.4 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 (de population) entre les différents groupes afin de voir quels groupes sont différents des autres. Dans le cas où par exemple l'interaction est significative, les tests 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 groupe différent 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 $\frac{m(m-1)}{2}$ comparaisons 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 Tukey(-Kramer) : Pour les tests de comparaison des m moyennes deux à deux, l'intervalle de confiance de Tukey(-Kramer) de niveau $1 - \alpha$ (Dans notre cas d'étude 95%) 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'autres méthodes qui ne seront pas abordées dans ce travail telles que : **Méthode de Scheffé, Méthode de Holm-Bonferroni, Correction de Bonferroni, etc.**

2.1.5 Taille d'effet

La taille d'effet dans un modèle d'analyse de variance est une mesure qui permet d'évaluer l'ampleur de la différence entre les groupes due à un effet. Pour le calcul de la taille d'effet, nous pouvons utiliser ω^2 (omega-squared) comme moyen de mesure. ω^2 est défini de la manière suivante :

$$\omega^2 = \frac{SCE_{facteur} - (k - 1) \cdot MS_{within}}{SCT + MS_{within}} \quad \text{avec} \quad MS_{within} = \frac{SCR}{N - k} \quad (11)$$

où :

- $SCE_{facteur}$ est la somme des carrés expliquée par le facteur en question
- SCT est la somme des carrés totale
- SCR est la somme des carrés résiduels
- k est le nombre de modalités du facteur
- N est le nombre total d'observations
- MS_{within} est la moyenne des carrés des erreurs au sein des groupes

Interprétation (Cohen 1988) : Il s'agit de la proportion de la variabilité de la variable dépendante qui peut être expliquée par le facteur en question. Une valeur de $\omega^2 = 0$ signifie qu'il n'y a aucune relation entre les deux, alors qu'une valeur de $\omega^2 = 1$ signifie que la relation est parfaite.

- Petite taille d'effet $\omega^2 \approx 0.01$
- Taille d'effet moyenne $\omega^2 \approx 0.06$
- Grande taille d'effet $\omega^2 \approx 0.14$

Contexte de l'étude et Problématique

L'étude inclut des collégiens de cycle 4 (5ème à la 3ème) issus de classes ordinaires ayant rendu les autorisations parentales, âgés entre 11 et 15 ans, dans des collèges allant de très défavorisés à très favorisés et souhaitant participer à l'étude. Un premier questionnaire est transmis aux élèves avant le début de l'étude afin de collecter les données personnelles de chaque participant : âge, taille, poids, fratrie, pratique d'activité physique (AP) et autres informations socio-culturelles. Ensuite, les élèves sont équipés d'accéléromètres (ActiGraph, modèle GT3X+) pour mesurer le niveau de MVPA (activité physique modérée à vigoureuse) chez les filles et les garçons pendant un cours d'EPS obligatoire de 2 heures. La collecte de données se déroule dans les collèges en Alsace et en Ile-de-France. Les écoles sont classées selon l'indice de Position sociale (catégorie d'IPS) : écoles défavorisées ($IPS < 89$), écoles moyennes ($90 \leq IPS \leq 114$) et écoles favorisées ($IPS > 115$). Au total, nous avons 462 observations et 28 variables. Le logiciel R (version 4.3.3) est utilisé pour faire toutes les analyses.

Voici un aperçu des premières lignes de mes données avec les colonnes utiles pour mon stage :

A tibble: 462 × 7

classe <chr>	mvp <dbl>	ecart_MVPA <dbl>	gender <chr>	cat_IPS <chr>	CA <dbl>	age <int>
3.P	23.833333	-5.282051...	F	moyen	3	15
3.P	32.666667	3.55128205	M	moyen	3	14
3.P	27.333333	-1.782051...	M	moyen	3	15
3.P	21.500000	-7.615384...	F	moyen	3	15
3.P	30.500000	1.38461538	M	moyen	3	15
3.P	28.000000	-1.115384...	F	moyen	3	15
3.P	28.500000	-0.615384...	M	moyen	3	15
3.P	32.500000	3.38461538	M	moyen	3	15
3.P	50.000000	20.88461...	M	moyen	3	15
3.P	28.333333	-0.782051...	F	moyen	3	15

1-10 of 462 rows Previous 1 2 3 4 5 6 ... 47 Next

FIGURE 1 – Extraits des données

L'objectif de notre étude est d'examiner les écarts de MVPA entre les filles et les garçons en tenant compte de trois variables : la nature de l'activité (CA = champs d'apprentissage), la catégorie socioculturelle de l'établissement (catégorie d'IPS = indice de position sociale) et le milieu géographique.

Partie Résultats

4.1 Analyse descriptive

La première partie de mon stage consiste à faire une analyse descriptive des données après les avoir nettoyées. Les statistiques descriptives sont présentées dans le tableau 4. Au total, nous avons eu 462 participants dans notre étude, dont 220 filles et 242 garçons. Les participants avaient en moyenne 13,65 ans. 47,61% des participants étaient des filles et 52,39% étaient des garçons. En termes de champs d'apprentissage, 12,13% étaient dans le champ 1, 26,15% dans le champ 2, 10,24% dans le champ 3 et environ 51,48% dans le champ 4. En ce qui concerne les catégories d'IPS, 31% étaient dans la catégorie basse, 22,37% dans la catégorie moyenne et 46,63% dans la catégorie haute. Pour les zones géographiques, 64,15% étaient en zones urbaines et 35,85% en zones rurales. La moyenne de MVPA pendant 2 heures pour les participants était de 35,14 minutes, avec 31,4 minutes pour les filles et 38,5 minutes pour les garçons. Les détails sur les catégories socio-professionnelles des parents sont présentés dans le tableau 4, ainsi que le nombre de participants dont les parents occupent l'une des professions listées. Le tableau 4 montre également la proportion de filles et de garçons dans différents champs d'apprentissage, catégories d'IPS et zones géographiques. En général, les proportions de filles et de garçons selon les différentes variables sont similaires, sauf pour certaines modalités de ces variables.

Variables	Total	Filles	Garçons
Participants	(n = 462)	(n = 220)	(n = 242)
Participants (%)	100%	47.61%	52.39%
Âge moyen	13.65	13.66	13.65
MVPA			
moyenne	35.14	31.4	38.5
médiane	34.83	31	37.67
écart-type	15.79	15.23	15.57
maximum	77.16	63.5	77.17
minimum	1.333	1.33	6
Catégories socio-professionnelles des parents (nombre)			
Agriculteurs	10	7	3
Artisans, commerçants, chefs d'entreprise	106	59	47
Autres inactifs	82	32	50
Cadres et professions intellectuelles supérieures	99	52	47
Employés	213	90	123
Ouvriers	66	28	38
Professions intermédiaires	116	59	57
Retraités	7	4	3
NA	43	23	20
Champs d'apprentissage			
Champ 1 (sports de performance)	12.13%	5.12%	7.01%
Champ 2 (sports de plein air)	26.15%	13.75%	12.40%
Champ 3 (activités artistiques)	10.24%	4.31%	5.93%
Champ 4 (activités d'opposition)	52.48%	24.53%	26.95%
Catégories socioculturelles			
Haute	46.63%	16.17%	14.83%
Basse	22.37%	11.86%	10.51%
Moyenne	31%	19.68%	26.95%
Zone géographique			
Rurale	35.85%	17.79%	18.06%
Urbaine	64.15%	29.92%	34.23%

TABLE 4 – Statistiques descriptives des participants

4.2 Modèles d'ANOVA

Au total, j'ai fait 3 modèles d'ANOVA trois à deux facteurs pour répondre aux problématiques de mon stage.

4.2.1 Examen de l'écart de MVPA entre Filles et Garçons selon le CA

On souhaite étudier ici l'effet du genre et du champ d'apprentissage sur le MVPA. Pour ce faire, nous utilisons un modèle d'ANOVA à deux facteurs (genre et CA) avec comme variable dépendante l'écart de MVPA à la moyenne de chaque classe. Ce choix de variable dépendante

étant imposé par l'étude et est justifié par le fait que certaines classes ont des temps d'activités différents et des enseignants différents, ce qui peut influencer le MVPA mesuré. Ce choix assure donc l'indépendance des observations et l'interprétation revient au même qu'à interpréter directement le MVPA. Dans la suite, 'ecart_MVPA' fera référence à l'écart de MVPA à la moyenne de chaque classe. Pour la contrainte sur le modèle, nous utilisons celle de cellule somme nulle. Le modèle s'écrit :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad (12)$$

$i = 1, \dots, 2$, $j = 1, \dots, 4$, $k = 1, \dots, K$, $\forall i, j, k$ $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$ avec un effet moyen général μ , α_i représente l'effet principal du genre, β_j représente l'effet principal du CA, un terme d'interaction γ_{ij} , les résidus ϵ_{ijk} avec k l'indice de répétition pour le couple (i, j) et σ^2 est la variance résiduelle.

Plan : On a le tableau de contingence suivant.

	CA 1	CA 2	CA 3	CA 4
F	20	53	49	98
M	26	51	54	111

TABLE 5 – Tableau des CA par genre

Le nombre d'observations diffère selon le croisement des modalités de CA et gender. Donc nos données sont déséquilibrées, on ne peut donc pas estimer de manière indépendante l'effet d'un facteur ou interactions des autres. Dans ce cas, nous utilisons l'ANOVA de type II car l'ordre des facteurs n'a pas d'importance dans mon étude. Avant de regarder les résultats, nous devons nous assurer que les hypothèses sont validées.

Hypothèses du modèle :

- **Indépendance** : L'hypothèse d'indépendance des observations est vérifiée. La liaison qui peut exister due à l'appartenance à une même classe et école est inhibée du fait du choix de l'écart à la moyenne de MVPA de chaque classe au lieu du MVPA.
- **Homoscédasticité** : Pour valider l'hypothèse d'homoscédasticité, nous visualisons le graphe des résidus en fonction des valeurs prédites.

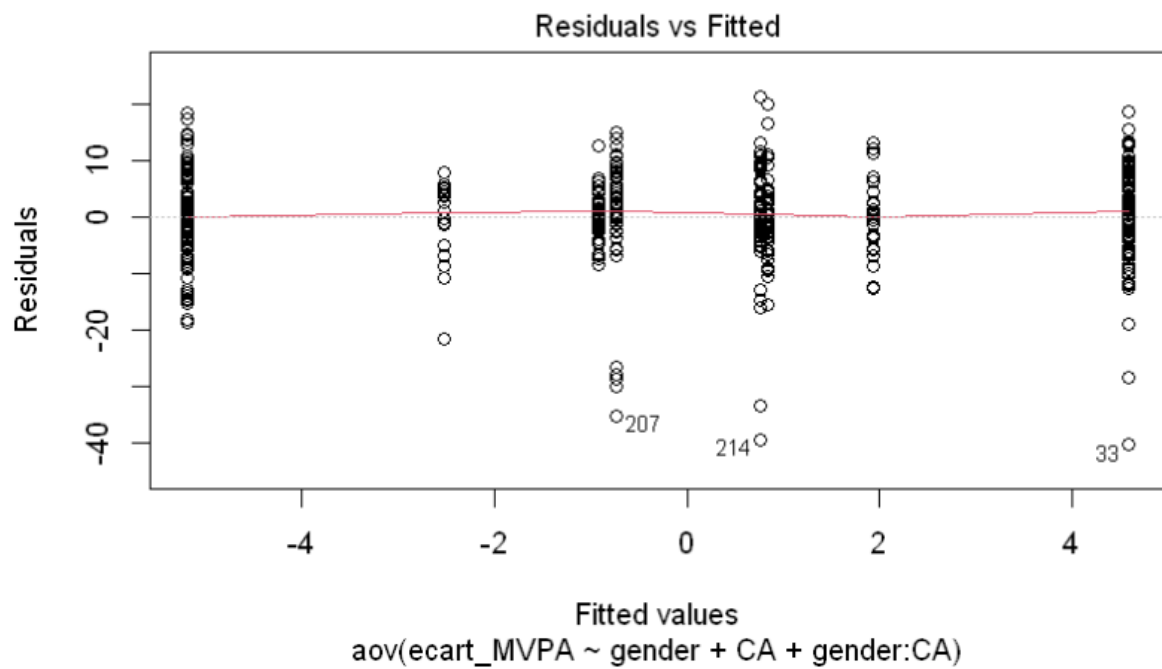


FIGURE 2 – Graphique de Diagnostic de l'hypothèse d'Homoscédasticité

Ce graphique nous permet de voir que la variance des résidus est similaire dans tous les groupes.

- **Normalité des résidus** : Visualisons le graphique du QQ plot.

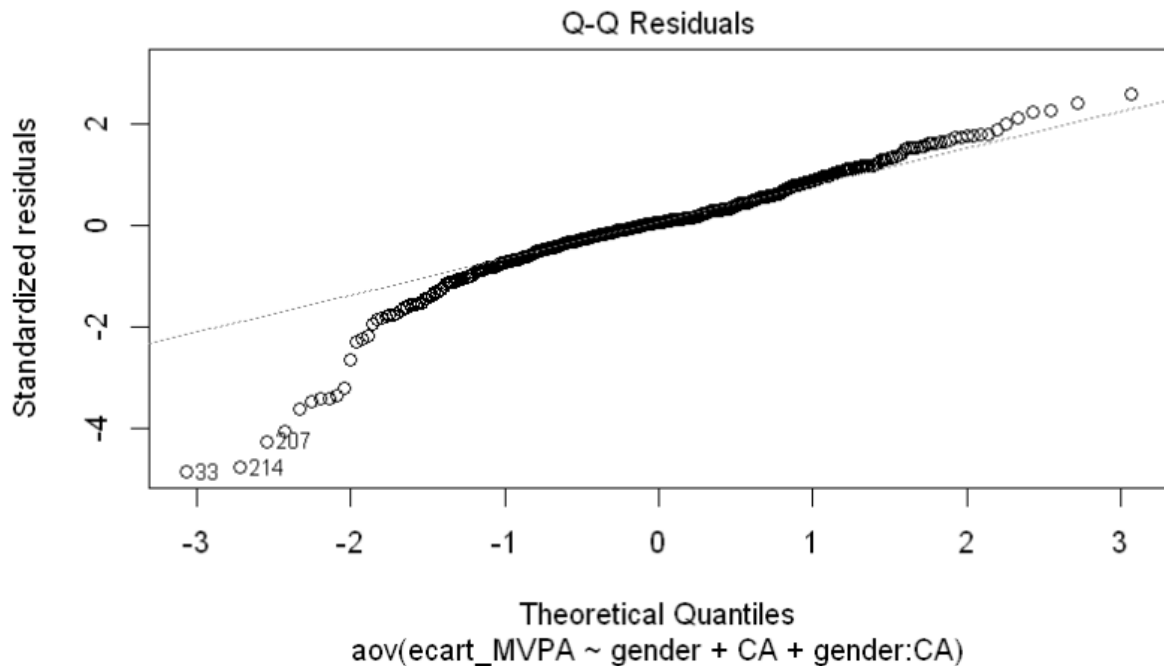


FIGURE 3 – QQ plot

Le graphique de quantiles contre quantiles permet de vérifier que l’hypothèse de normalité des résidus est raisonnable.

Résultats : Test du modèle complet : On veut tester les hypothèses

- $H_0 : \alpha_i = \alpha_{i'} \text{ et } \beta_j = \beta_{j'} \text{ et } \gamma_{ij} = \gamma_{i'j'} \quad \forall i' \neq i, \quad j' \neq j$
- $H_1 : \alpha_i \neq \alpha_{i'} \text{ ou } \beta_j \neq \beta_{j'} \text{ ou } \gamma_{ij} \neq \gamma_{i'j'} \quad \forall i' \neq i, \quad j' \neq j$

On a le tableau suivant :

Res.Df	RSS	Df	Sum of Sq	F value	Pr(>F)
461	36916				
454	31593	7	5323.2	10.928	8.944e-13 ***

TABLE 6 – Tableau des résultats de l’ANOVA

La statistique de Fisher $F = \frac{\frac{SCE}{(IJ-1)}}{\frac{SCR}{(n-IJ)}} = 10.928$ (la variabilité expliquée par le genre et CA est 11 fois supérieure à la variabilité résiduelle), sous H_0 , $F \sim \mathcal{F}_{7,454}$. Ce qui correspond à une probabilité critique de 8.944×10^{-13} , donc on rejette H_0 . On peut remarquer que l’ajustement du modèle n’est pas très satisfaisant. En effet, on a $R^2 = 1 - \frac{31593}{36916} = 0.144$.

Test des différents effets :

On a le tableau suivant :

Source	Sum Sq	Df	F value	Pr(>F)
gender	3585.2	1	51.5200	2.911e-12 ***
CA	6.4	3	0.0307	0.9927
gender :CA	1738.0	3	8.3252	2.122e-05 ***
Résidus	31592.8	454		

TABLE 7 – Effet des différents facteurs (type II) dans le modèle d’analyse de variance à deux facteurs avec interaction

Effet de l’interaction :

Les hypothèses testées sont :

- H_0 : Il n’y a pas de différence significative entre le modèle complet et le modèle ne contenant pas l’interaction.
- H_1 : Il y a une différence significative entre le modèle complet et le modèle ne contenant pas l’interaction.

La statistique $F_{gender*CA} = 8.3252$, qui sous H_0 suis une loi de Fisher à 3 et 454 ddl. La probabilité critique associée au test vaut 2.122×10^{-5} . L’effet de l’interaction entre genre et CA est significatif au niveau 0.05. Donc nous n’allons plus accorder trop d’importance au test sur les facteurs principaux, les 2 facteurs sont influents via leur interaction.

Comparaison des ecart_MVPA :

Nous avons détecter un effet significatif de l’interaction et ici nous sommes donc intéressé par la différence de ecart_MVPA entre les filles et garçons selon les différentes CA. On veut tester donc les hypothèses :

- $H_0 : \theta_{i_1j} = \theta_{i_2j}$
- $H_1 : \theta_{i_1j} \neq \theta_{i_2j}$

où $\theta_{i_1j_1}$ et $\theta_{i_2j_2}$ sont les moyennes dans chaque groupe différent et $i_1, i_2 \in \{1, 2\}$ avec $i_1 \neq i_2$; $j \in \{1, \dots, 4\}$. Le tableau 8 présente les test post hoc de tukey-kramer, comparaison des moyennes ecart_MVPA des filles et garçons selon le champs d’apprentissage (CA)

	CA1	CA2	CA3	CA4
Écart moyen entre fille et garçon des écarts à la moyenne de MVPA de chaque classe	-4.4548	-1.4898	-1.7470	-9.7647
Intervalle de confiance	-12.010 à 3.100	-6.4730 à 3.493	-6.759 à 3.265	-13.286 à -6.244
P-valeur	0.6235	0.9850	0.96	0.000

TABLE 8 – Test de comparaison des moyennes de ecart_MVPA des filles et garçon dans les CA différents

Le MVPA (ou ecart_MVPA car ils ont même interprétation) des garçons est significative-ment supérieur à celui des filles dans le champ 4 ($p_valeur < .05$, on rejette H_0) par contre dans les autres champs cette différence n’est pas significative avec des p_valeur largement supérieur à 5% (on conserve H_0).

Le graphe ci-dessous résume cette situation :

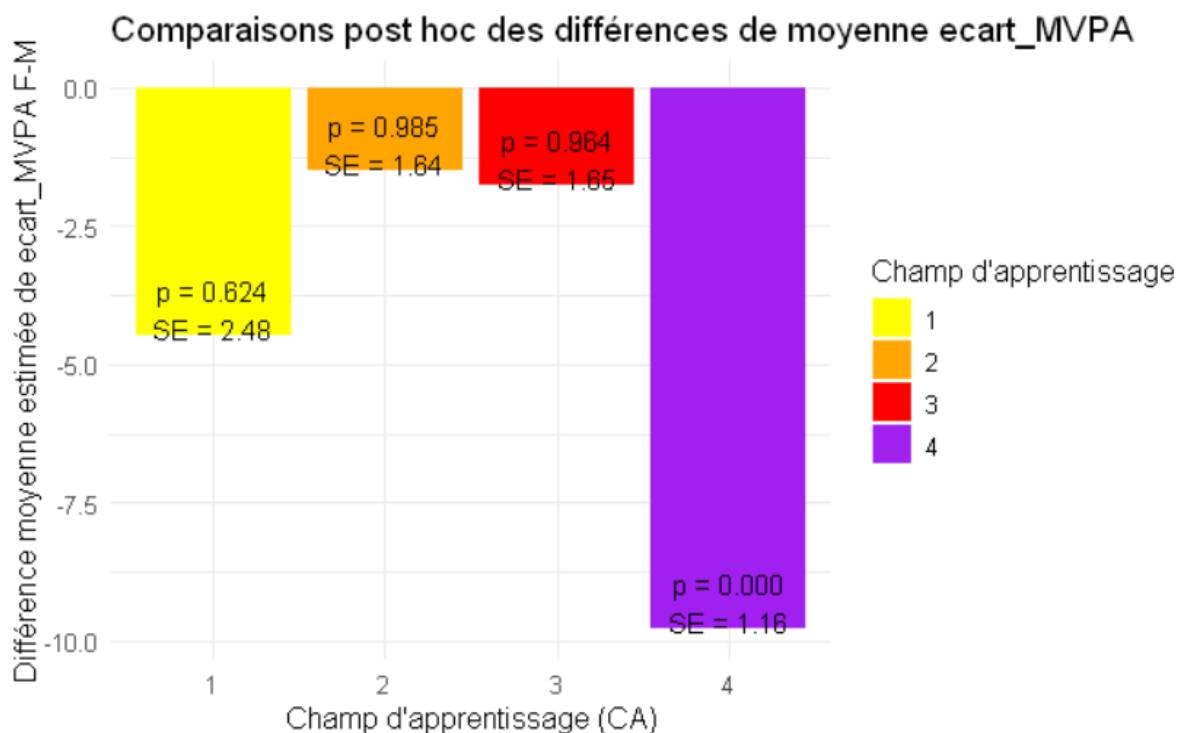


FIGURE 4 – Diagramme en barre représentatif des test de comparaison des moyennes de ecart_MVPA des filles et garçon dans les CA différents avec les p-valeur et écart_type

On voit clairement sur ce graphe que dans le champ 4 les garçons s’impliquent plus que les filles ce qui n’est clairement pas le cas dans les autres champs d’apprentissage.

Taille d’effet :

Dans le champ 4 il y a une différence significative de moyenne ecart_MVPA entre les filles et garçons ce qui est traduit par un effet significatif de l’interaction entre le genre et CA. La question qu’on se pose est de savoir quelle est la force de cet effet.

On a les résultats suivants :

Parameter	ω^2	95% CI
gender	0.10	[0.06, 1.00]
CA	0.00	[0.00, 1.00]
gender :CA	0.05	[0.02, 1.00]

TABLE 9 – Tableau des tailles d’effet (ω^2) et intervalle de confiance à 95% (unilatéral)

La taille d’effet pour l’interaction entre genre et CA est 0.05 ce qui signifie que 5% de la variance totale de la variable dépendante (ecart_mvpa) peut être attribuée à l’effet de l’interaction. Dans le métier cette taille d’effet est modérément importante ce qui nous permet de tenir compte de ce résultat dans la pratique (écart de mvpa entre les filles et garçons observé dans le champ 4 est à prendre en compte dans la pratique car jugé important).

4.2.2 Examen de l'écart de MVPA entre Filles et Garçons

L'objectif ici est de voir si l'étude fait produit un résultat similaire à ceux des travaux antérieurs.

4.2.3 Examen de l'écart de MVPA entre Filles et Garçons selon la catégorie IPS

L'objectif ici est d'examiner l'effet du genre et de la catégorie IPS sur le MVPA. Nous utilisons donc un modèle d'ANOVA à deux facteurs (genre et IPS catégorie) avec comme variable dépendante `ecart_MVPA`. Pour la contrainte sur le modèle, nous utilisons celle de somme des coefficients nulle. Le modèle s'écrit :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad (13)$$

$i = 1, \dots, 2$, $j = 1, \dots, 3$, $k = 1, \dots, K$, $\forall i, j, k$ $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$ avec un effet moyen général μ , α_i représente l'effet principal du genre, β_j représente l'effet principal de l'IPS catégorie, un terme d'interaction γ_{ij} , les résidus ϵ_{ijk} avec k l'indice de répétition pour le couple (i, j) et σ^2 est la variance résiduelle.

Plan : On a le tableau de contingence suivant

	Élevée	Faible	Moyenne
F	93	48	79
M	89	44	109

TABLE 10 – Tableau de contingence entre le genre et l'IPS catégorie

On voit clairement que nos données sont déséquilibrées. Nous utilisons donc une ANOVA de Type II. En effet, l'ordre d'insertion des facteurs dans le modèle n'est pas important dans le cadre de mon étude. Avant de regarder les résultats, nous devons nous assurer que les hypothèses du modèle sont validées.

Hypothèses du modèle :

- **Indépendance** : L'hypothèse d'indépendance des observations est vérifiée.
- **Homoscédasticité** : Pour valider l'hypothèse d'homoscédasticité, nous visualisons le graphe des résidus en fonction des valeurs prédites.

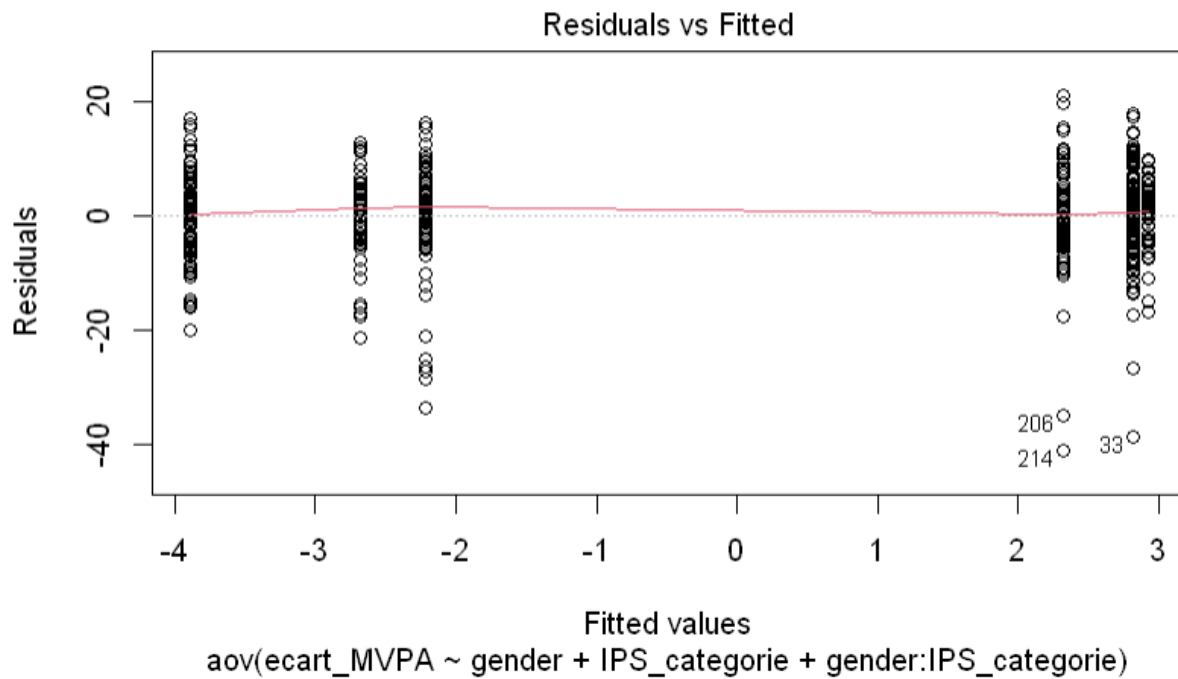


FIGURE 5 – Graphique de Diagnostic de l'hypothèse d'Homoscédasticité

Ce graphique nous permet de voir que la variance des résidus est similaire dans tous les groupes.

— **Normalité des résidus** : Visualisons le graphique du QQ plot.

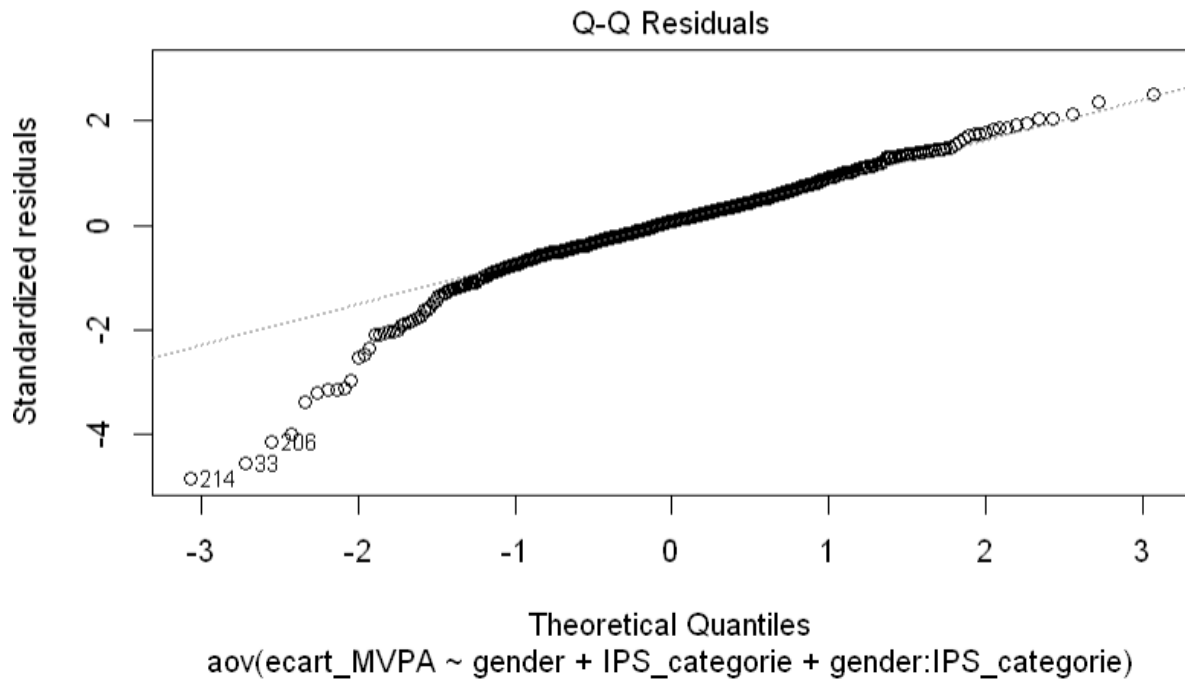


FIGURE 6 – QQ plot

Le graphique de quantiles contre quantiles permet de vérifier que l’hypothèse de normalité des résidus est raisonnable.

Résultats :

Test du modèle complet :

On veut tester les hypothèses

- $H_0 : \alpha_i = 0$ et $\beta_j = 0$ et $\gamma_{ij} = 0 \quad \forall i, j$
- $H_1 : \alpha_i \neq 0$ ou $\beta_j \neq 0$ ou $\gamma_{ij} \neq 0 \quad \forall i, j$

On a le tableau suivant :

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
461	36916				
456	33199	5	3717.1	10.211	2.74e-09 ***

TABLE 11 – Tableau des résultats de l’analyse de variance

La statistique de Fisher $F = \frac{\frac{SCE}{(IJ-1)}}{\frac{SCR}{(n-IJ)}} = 10.211$ (la variabilité expliquée par le genre et L’IPS catégorie est 10 fois supérieure à la variabilité résiduelle), sous H_0 , $F \sim \mathcal{F}_{5,456}$. Ce qui correspond à une probabilité critique de 2.74×10^{-09} , donc on rejette H_0 . On peut remarquer que l’ajustement du modèle n’est pas très satisfaisant. En effet, on a $R^2 = 1 - \frac{33199}{36916} = 0.1006$.

Test des différents effets :

On a le tableau suivant :

Source	Sum Sq	Df	F value	Pr(>F)
gender	3610	1	49.5860	7.028e-12 ***
IPS_categorie	31	2	0.2152	0.8064
gender :IPS_categorie	107	2	0.7349	0.4801
Residuals	33199	456		

TABLE 12 – Tableau des résultats de l’analyse de variance pour l’effet du genre et de la catégorie IPS

Effet de l’interaction

Ici nous testons les hypothèses

- $H_0 : Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad \forall i, j, k$
- $H_1 : Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad \forall i, j, k$

La statistique de test $F_{gender*IPS_categorie} = 0.7349$, qui sous H_0 suit une loi de Fisher à 2 et 456 ddl. La p-valeur associé au test vaut 0.4801 donc l’hypothèse de non interaction H_0 est conservée.

Effet de l’IPS_categorie

Nous souhaitons savoir si l’IPS_categorie possède un effet sur ecart_MVPA. Nous testons alors les hypothèses

- $H_0 : Y_{ijk} = \mu + \alpha_i + \epsilon_{ijk} \quad \forall i, j, k$
- $H_1 : Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad \forall i, j, k$

La statistique de test $F_{IPS_categorie} = 0.2152$ qui sous H_0 suit une loi de Fisher à 2 et 456 ddl. La p-valeur observé vaut 0.8064 donc nous conservons H_0 , il n’existe pas d’effet de l’IPS_categorie sur ecart_MVPA.

Donc l’écart de MVPA entre fille et garçons ne dépendent pas de la catégorie socio-professionnelle.

4.2.4 Examen de l’écart de MVPA entre Filles et Garçons selon le milieu géographique

Afin de savoir si les variables genre et milieu géographique ont un effet sur le MVPA, nous allons utiliser une ANOVA à deux facteurs (genre et géographie) avec comme variable dépendante ecart_MVPA. Pour la contrainte sur le modèle, nous utilisons celle de somme des coefficients nulle.

Le modèle s’écrit :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad (14)$$

$i = 1, \dots, 2, \quad j = 1, \dots, 4, \quad k = 1, \dots, K, \quad \forall i, j, k \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$ avec un effet moyen général μ , α_i représente l’effet principal du genre, β_j représente l’effet principal du milieu géographique, un terme d’interaction γ_{ij} , les résidus ϵ_{ijk} avec k l’indice de répétition pour le couple (i, j) et σ^2 est la variance résiduelle.

Plan :

Nous obtenons le tableau de contingence suivant

	Rural	Urbain
F	101	119
M	102	140

TABLE 13 – Tableau de contingence

Les données sont déséquilibrées, nous utiliserons donc une ANOVA de type II (l'ordre d'insertions des facteurs dans le modèle n'a pas d'importance). Avant de regarder les résultats, nous devons nous assurer que les hypothèses du modèle sont validées.

Hypothèses du modèle :

Indépendance : L'hypothèse d'indépendance des observations est vérifiée.

Homoscédasticité : Visualisons le graphe des résidus en fonction des valeurs prédites.

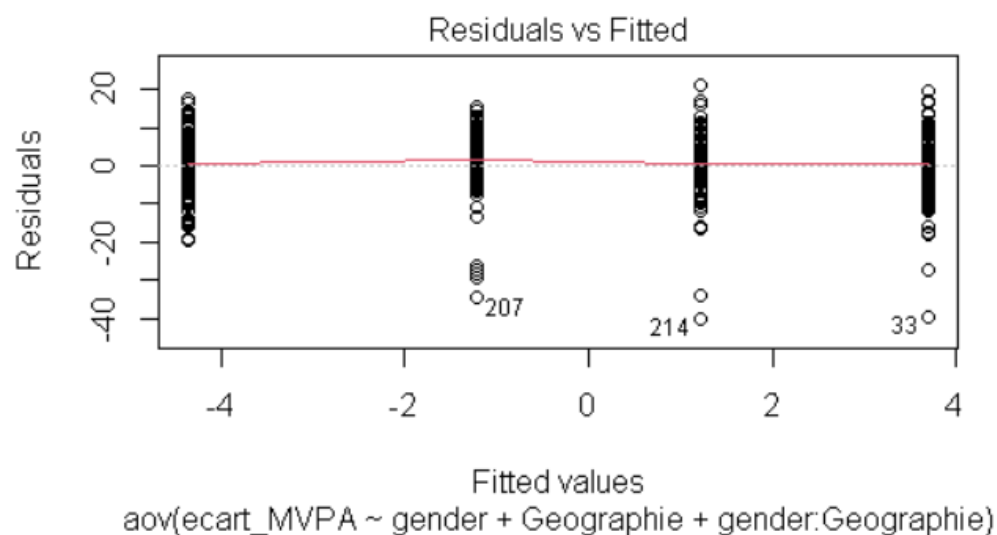


FIGURE 7 – Graphique de Diagnostic de l'hypothèse d'Homoscédasticité

Ce graphique nous permet de voir que la variance des résidus est similaire dans tous les groupes.

Normalité des résidus : Visualisons le graphique du QQ-plot.

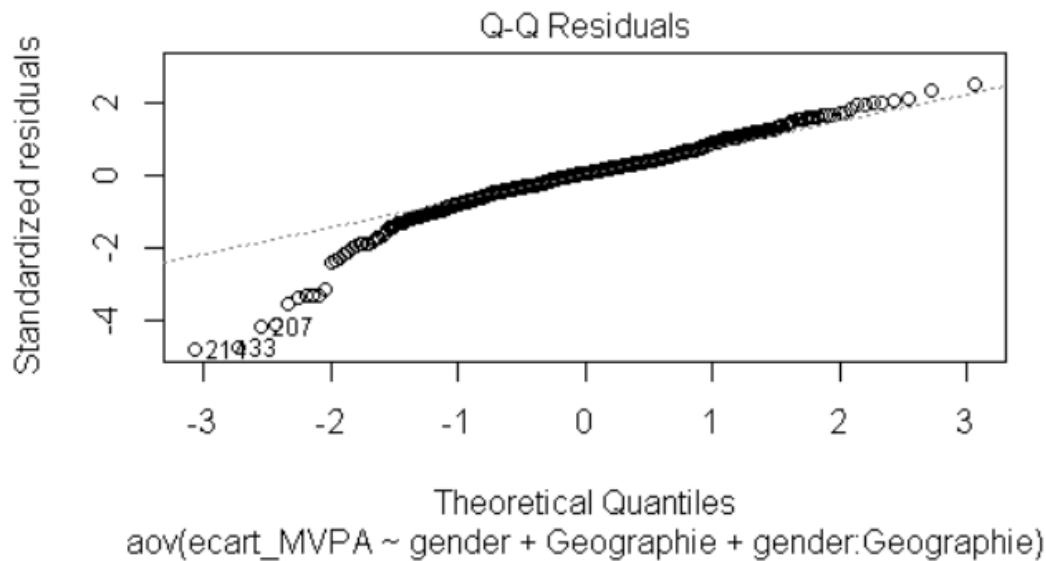


FIGURE 8 – QQ plot

Le graphique de quantile contre quantile permet de vérifier que l'hypothèse de normalité des résidus est raisonnable.

Résultats :

Test du modèle complet :

On souhaite tester

- $H_0 : \alpha_i = 0 \text{ et } \beta_j = 0 \text{ et } \gamma_{ij} = 0 \quad \forall i, j$
- $H_1 : \alpha_i = 0 \text{ ou } \beta_j = 0 \text{ ou } \gamma_{ij} = 0 \quad \forall i, j$

Voici les résultats :

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
461	36916				
458	32439	3	4477.3	21.072	8.375e-13 ***

TABLE 14 – Tableau des résultats de l'analyse de variance

La statistique de test $F = \frac{\frac{SCE}{(IJ-1)}}{\frac{SCR}{(n-IJ)}} = 21.072$ (la variabilité expliquée par le genre et le milieu géographique est 21 fois supérieure à la variabilité résiduelle), sous H_0 , $F \sim \mathcal{F}_{3,458}$. Ce qui correspond à une probabilité critique de 8.375×10^{-13} , donc on rejette H_0 . On peut remarquer que l'ajustement du modèle n'est pas très satisfaisant. En effet, on a $R^2 = 1 - \frac{32439}{36916} = 0.1212$.

Test des différents effets :

On obtient le tableau ci-dessous :

	Sum Sq	Df	F value	Pr(>F)
gender	3584	1	50.6006	4.386e-12
Geographie	5	1	0.0725	0.7879208
gender :Geographie	893	1	12.6145	0.0004224
Residuals	32439	458		

TABLE 15 – Analysis of Variance Table

Effet de l'interaction

On souhaite tester :

- $H_0 : Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad \forall i, j, k$
- $H_1 : Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad \forall i, j, k$

La statistique de test $F_{gender*geographie} = 12.6145$, qui sous H_0 suit une loi de Fisher à 1 et 458 ddl. La p-valeur associé au test vaut 0.0004224 donc l'effet de l'interaction entre genre et milieu géographique est significatif au niveau de 5%. Les deux facteurs sont influents via leur interaction, il n'est donc pas nécessaire de tester leur influence respective.

Comparaison des écarts_MVPA :

Un effet significatif de l'interaction est détecter et ici nous sommes intéressé par la différence de ecart_MVPA entre les filles et garçons selon les différents milieu géographique. On veut tester les hypothèses :

- $H_0 : \theta_{i_1j} = \theta_{i_2j}$
- $H_1 : \theta_{i_1j} \neq \theta_{i_2j}$

où $\theta_{i_1j_1}$ et $\theta_{i_2j_2}$ sont les moyennes dans chaque groupe différent et $i_1, i_2 \in \{1, 2\}$ avec $i_1 \neq i_2$; $j \in \{1, 2\}$. Le tableau suivant présente la comparaison des ecart_MVPA des filles et garçons selon le champs d'apprentissage

	Rural	Urbain
Écart entre filles et garçons des écarts à la moyenne de MVPA de chaque classe	-2.44	-8.06
Intervalle de confiance	-5.49 à 0.602	-10.761 à -5.350
P-valeur	0.1650	0.000

TABLE 16 – Comparaison des écarts à la moyenne de MVPA par genre et par localisation

Le MVPA (ou ecart_MVPA) est significativement supérieur à celui des filles dans le milieu urbain (p-valeur $< .05$, on rejette H_0) par contre dans le milieu rural cette différence n'est pas significative (p-valeur $> .05$, on conserve H_0).

Le graphe ci-dessous résume cette situation :

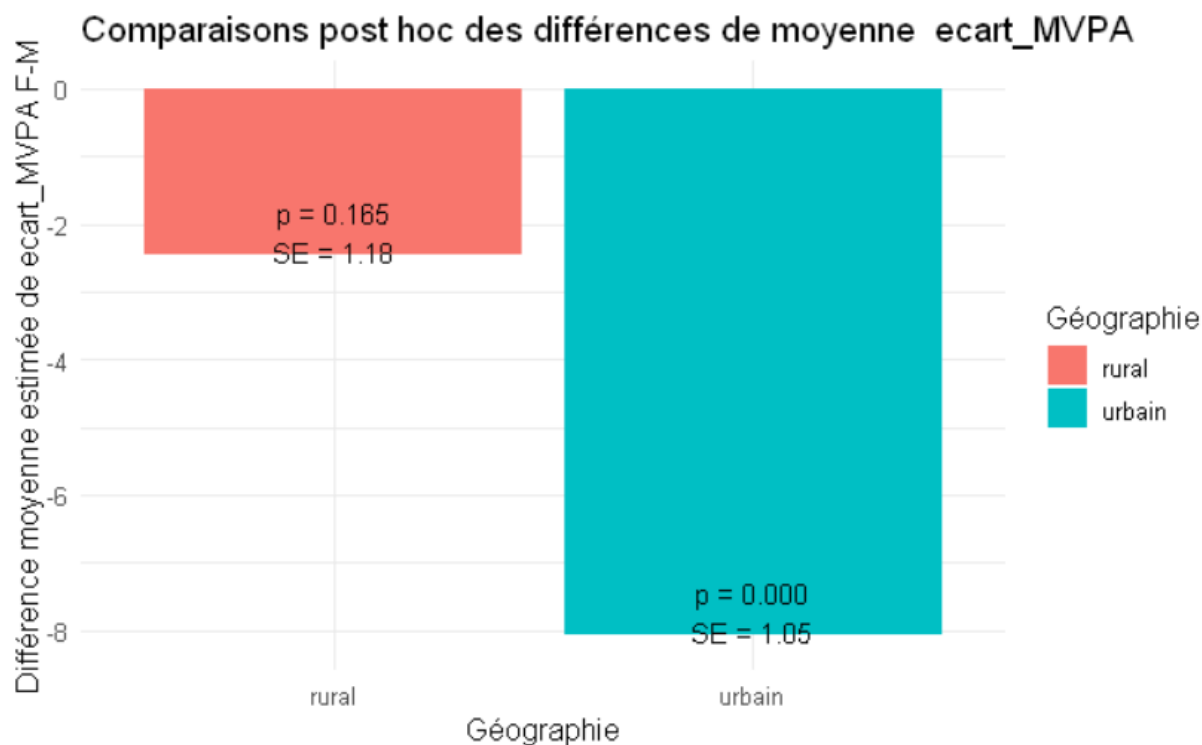


FIGURE 9 – Diagramme en barre représentatif des test de comparaison des moyennes de ecart_MVPA des filles et garçon dans les CA différents avec les p-valeur et écart_type

On voit clairement sur ce graphe que dans le milieu urbain les garçons s'impliquent plus que les filles ce qui n'est pas clairement le cas en milieu rural.

Taille d'effet : En milieu urbain il y a une différence significative de moyenne ecart_MVPA entre fille et garçon ce qui est traduit par un effet significatif de l'interaction entre le genre et le milieu géographique. Ici la question qu'on se pose est de savoir quelle est la force de ce effet. On a le tableau suivant :

Parameter	ω^2	95% CI
gender	0.10	[0.06, 1.00]
Geographie	0.00	[0.00, 1.00]
gender :Geographie	0.02	[0.01, 1.00]

TABLE 17 – Taille d'effet ω^2 et intervalle de confiance à 95% (unilatéral)

La taille d'effet pour l'interaction entre genre et milieu géographique est 0.02 ce qui signifie que 5% de la variance totale de la variable dépendante (ecart_MVPA) peu être attribuée à l'effet de l'interaction. Dans le métier cette taille d'effet est faible. La différence de ecart_MVPA entre fille et garçon observé en milieu urbain existe belle et bien mais cette doifférence n'est pas assez important pour en tenir compte dans la pratique.

Discussion sur les limites des résultats

Il est à noter que le calcul et l'interprétation des résultats de la taille d'effet est dans le cas de plan non équilibré un peu plus complexe du au fait qu'une partie de la variance est manquante. En effet comme la variabilité totale ne peut pas être décomposé en la variabilité des différents effet donc certaine variabilité qui ne peuvent pas être attribué à un effet spécifique disparaît dans le cas de l'ANOVA de type II ou III. Donc même si la taille d'effet n'est pas assez précise dans le cas de plan non équilibré, il nous donne quand même une vision globale sur la force de l'effet du facteur (plan pas trop déséquilibré).

Conclusion