

Présentation de l'étude de la relation entre les performances d'un étudiant et des facteurs environnementaux et sociaux

Nom : René Picard

Date : 25 avril 2024

Introduction

- Problématique : Trouver les facteurs qui influence la performance des étudiants québécois au secondaire
- Les données ont été récolté par l'université de Camerino

Objectif

Trouver les facteurs et les combinaisons de facteurs qui sont significatif dans la réussite étudiante.

Méthodologie

Variable réponse

- Nous avons gardé les 2 variables réponses de la partie 1 et ajouté 2 autres variables réponses.
 - La note finale en mathématique.
 - La note finale au cours de portugais.
- La différence de note au cours de mathématique.
 - La différence de note au cours de portugais.

Variables prédictrices

- Nous avons choisis les mêmes que pour la partie 1.
- Temps de voyage.
- Temps d'étude.
- Support scolaire.
- Support familial.
- Cours privés.
- Internet.
- Consommation d'alcool les jours d'école
- Consommation d'alcool les fins de semaine.

Hypothèses

- Hypothèse nulle : Les moyennes des notes finales (et respectivement la différence des notes) pour le cours de mathématiques (et respectivement de portugais) sont identiques.
- Hypothèse alternative : Au moins une des moyennes des notes finales (ou respectivement la différence des notes) pour le cours de mathématiques (ou respectivement de portugais) est différente.

Technique d'analyse

- Risque α fixé à 5 %.
- Création d'un modèle incluant toutes les interactions possibles (deux par deux) des variables prédictives.
- Vérification des trois postulats pour l'ANOVA.
- Réalisation d'une ANOVA.

Résultats

Premier modèle

$$Y_{1m} = b_0 + \sum_{i=1}^8 b_i X_i + \sum_{i=1}^7 \sum_{j=i+1}^8 b_{ij} (X_i \cdot X_j) + \text{epsilon}$$

où

Y_{1m} = Note Finale au cours de mathématique ,

X_1 = temps d ' étude ,

X_2 = temps de voyage ,

X_3 = support scolaire ,

X_4 = support familial ,

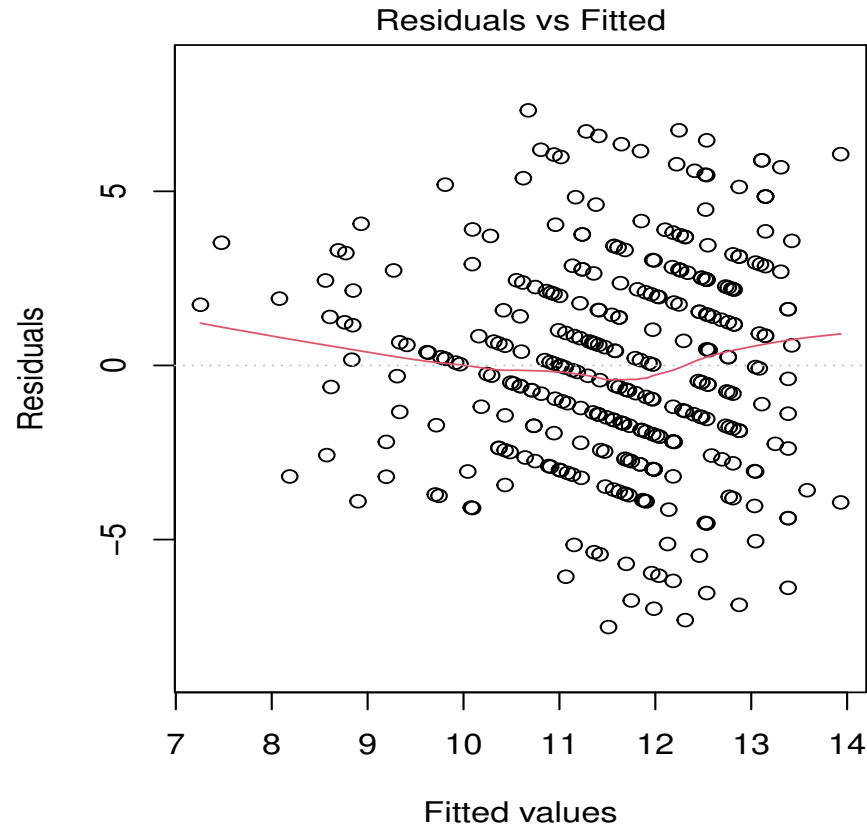
X_5 = cours privés ,

X_6 = internet ,

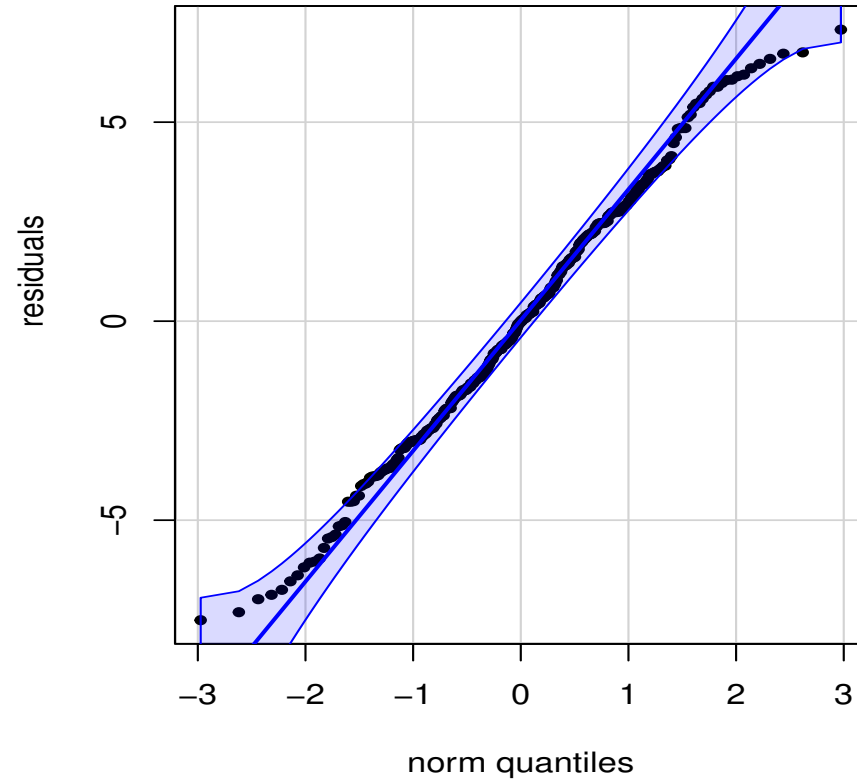
X_7 = Consommation d ' alcool les jours d ' école ,

X_8 = Consommation d ' alcool les fin de semaine

Homoscédasticité



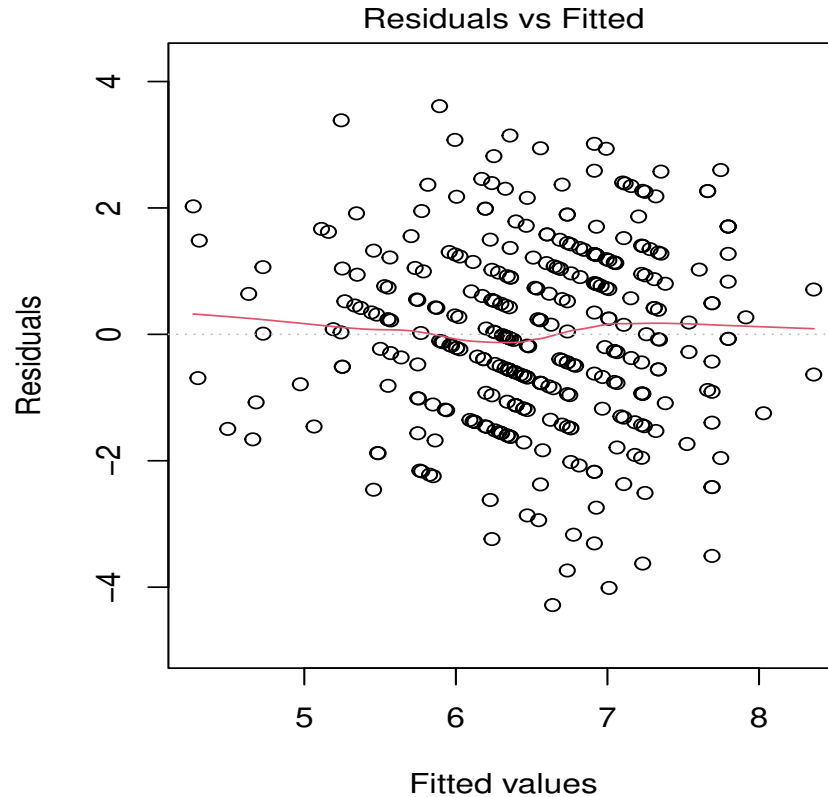
Normalité des résidus



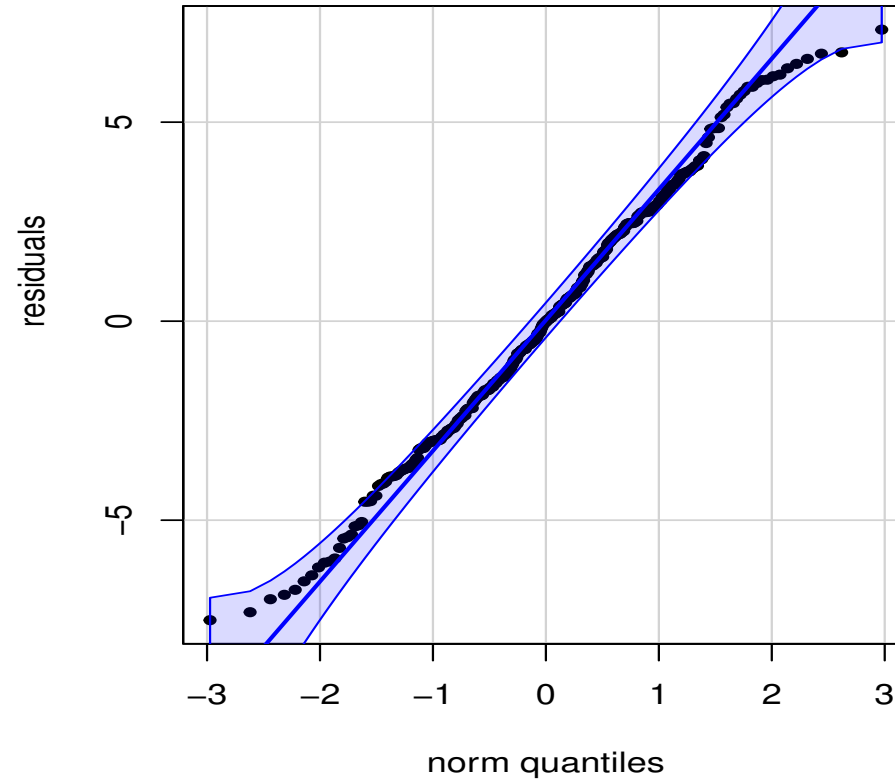
Normalité des résidus

- Test de Shapiro
- $W = 0.99339$, p-valeur = 0.1417
- Donc transformation de boxcox
- Création d'un nouveau modèle

Homoscédasticité



Normalité des résidus



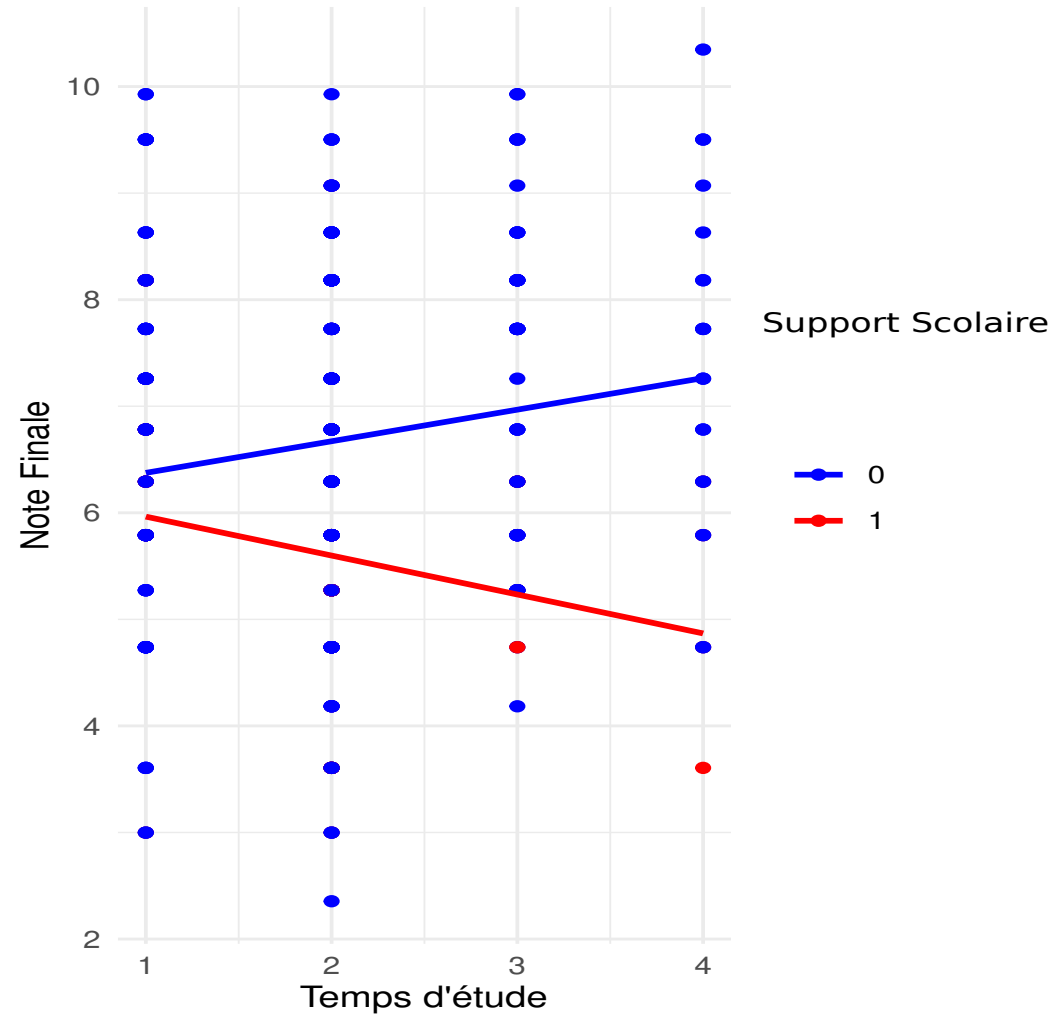
Normalité des résidus et égalité des variances

- Test de Shapiro.
- $W = 0.99328$, p-valeur = 0.1335
- Test de Durbin-Watson.
- $DW = 1.9408$, p-valeur = 0.2859

ANOVA

| Variable 1 | Variable 2 | P_Valeur |
|-------------------|-------------------|-----------------|
| Temps d'étude | Support Scolaire | 0.014286 |
| Temps d'étude | Support Familial | 0.033992 |

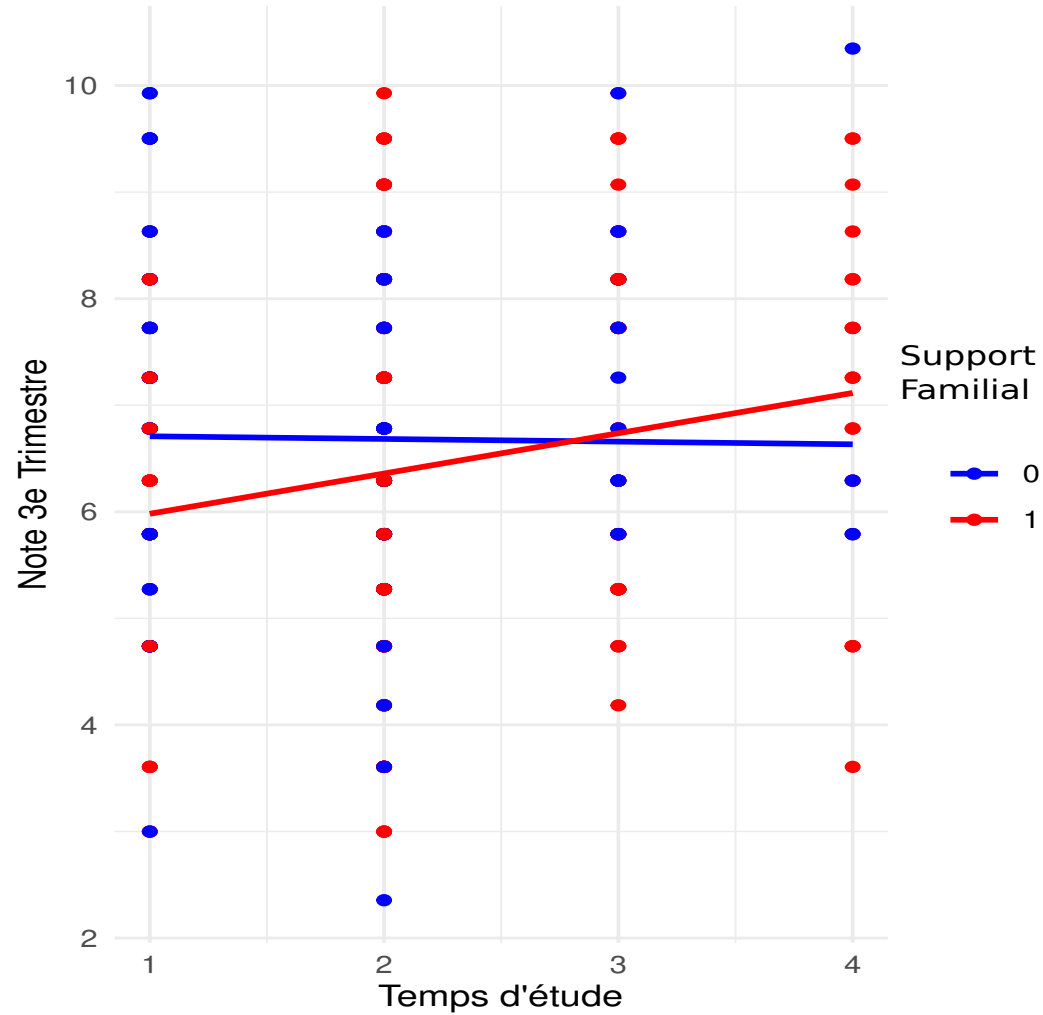
Interaction entre temps d'étude et support scolaire avec la note finale



Interaction Temps d'Étude et Support Scolaire

- Lorsque les élèves ne reçoivent pas de support scolaire, on observe que le temps consacré à l'étude améliore la note finale.
- Toutefois, cet effet positif du temps d'étude semble s'inverser en présence d'un support scolaire.
- Ces résultats sont ambivalents mais pas nécessairement illogiques.

Interaction entre temps d'étude et le support familial avec la note finale



Interaction Temps d'Étude et Support Familial

- Lorsque les étudiant ne reçoivent pas de support familial, on observe que le temps consacré à l'étude n'améliore pas la note finale.
- Lorsque les étudiants bénéficient d'un soutien familial, le temps d'étude s'avère bénéfique uniquement aux niveaux 3 et 4.

Deuxième modèle

$$Y_{1d} = b_0 + \sum_{i=1}^8 b_i X_i + \sum_{i=1}^7 \sum_{j=i+1}^8 b_{ij} (X_i \cdot X_j) + \text{epsilon}$$

où

Y_{1d} = Différence de note au cours de mathématique ,

X_1 = temps d ' étude ,

X_2 = temps de voyage ,

X_3 = support scolaire ,

X_4 = support familial ,

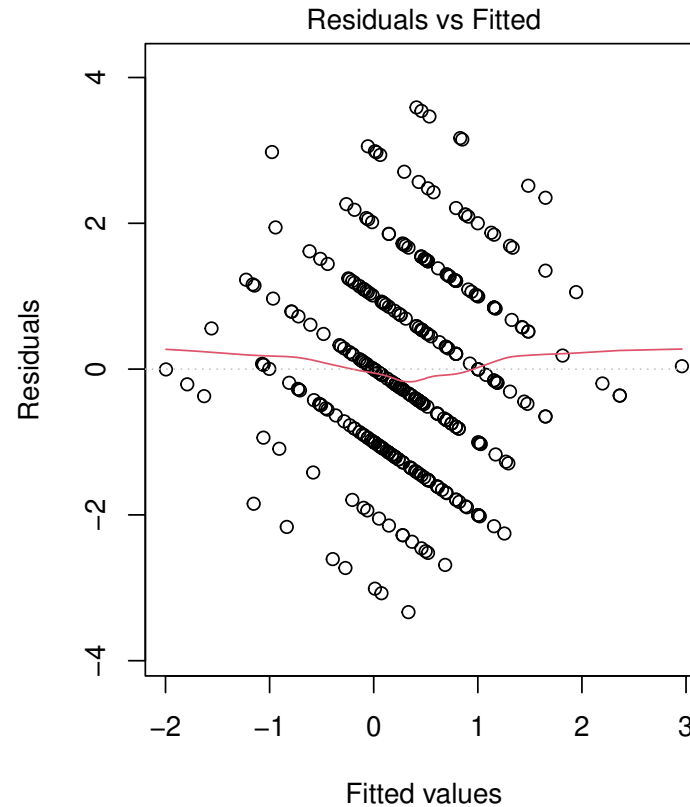
X_5 = cours privés ,

X_6 = internet ,

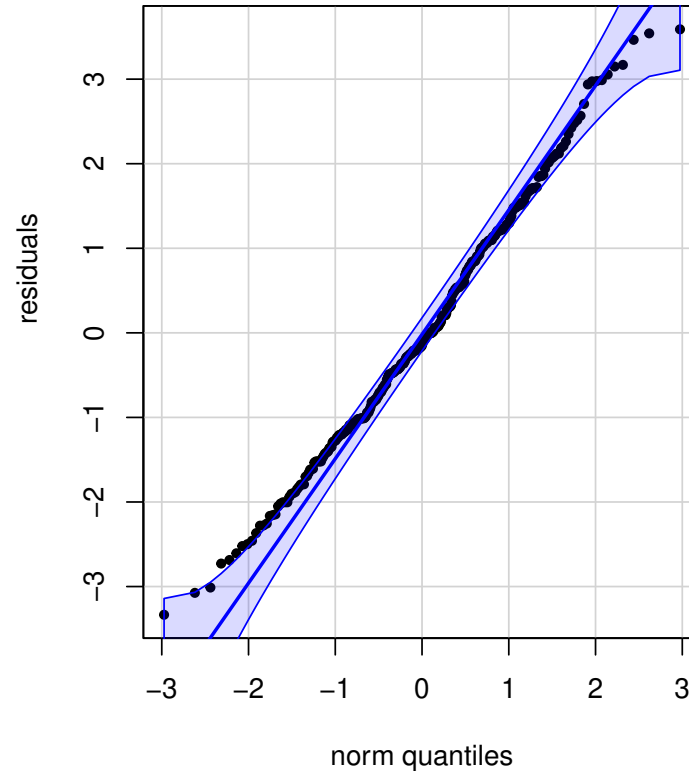
X_7 = Consommation d ' alcool les jours d ' école ,

X_8 = Consommation d ' alcool les fin de semaine

Homoscédasticité



Normalité des résidus



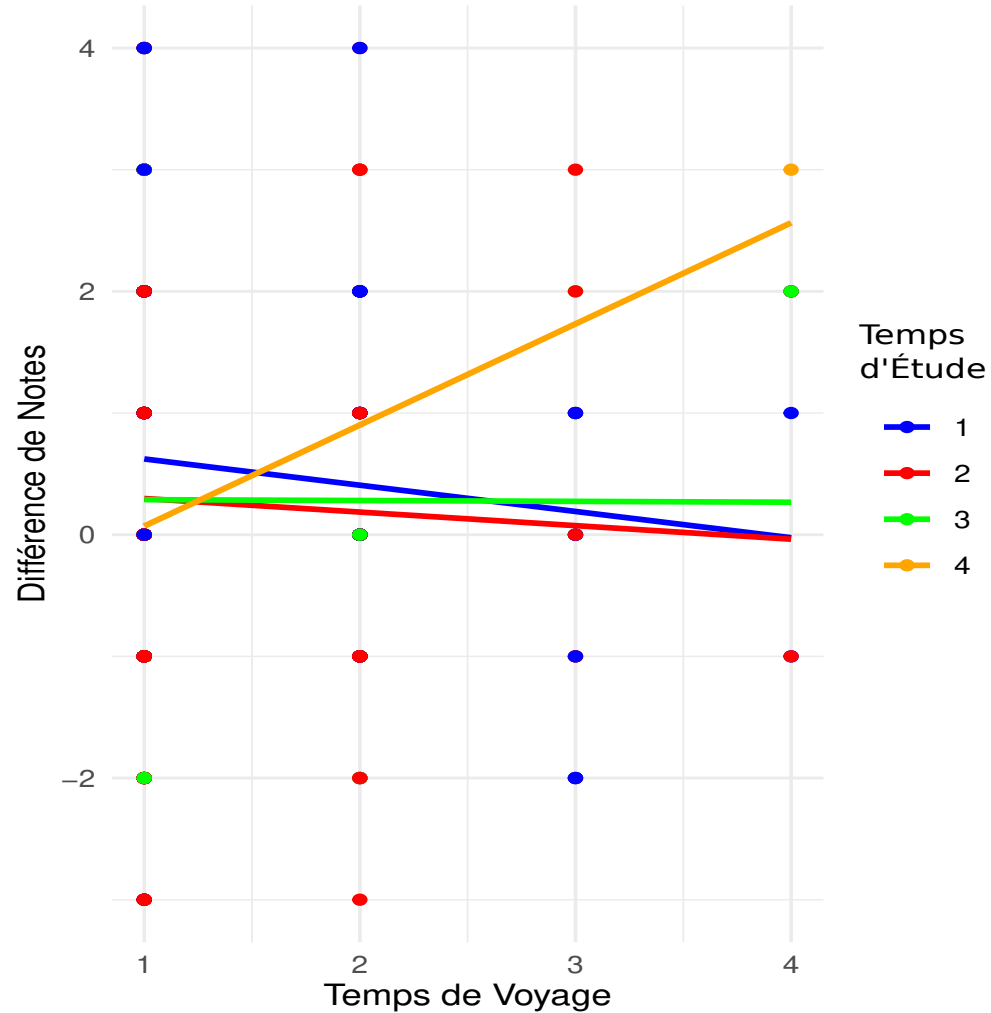
Normalité des résidus et égalité des variances

- Test de Shapiro.
- $W = 0.99278$, p-valeur = 0.1004
- Test de Durbin-Watson.
- $DW = 1.8657$, p-valeur = 0.0996

ANOVA

| Variable 1 | Variable 2 | P_Valeur |
|------------------|--------------------------------------|----------|
| Temps de voyage | Temps d'étude | 0.002244 |
| Temps de voyage | Support Familial | 0.001524 |
| Temps de voyage | Consommation d'Alcool Fin de Semaine | 0.020025 |
| Support Scolaire | Support Familial | 0.029212 |
| Support Scolaire | Consommation d'alcool Semaine | 0.001957 |

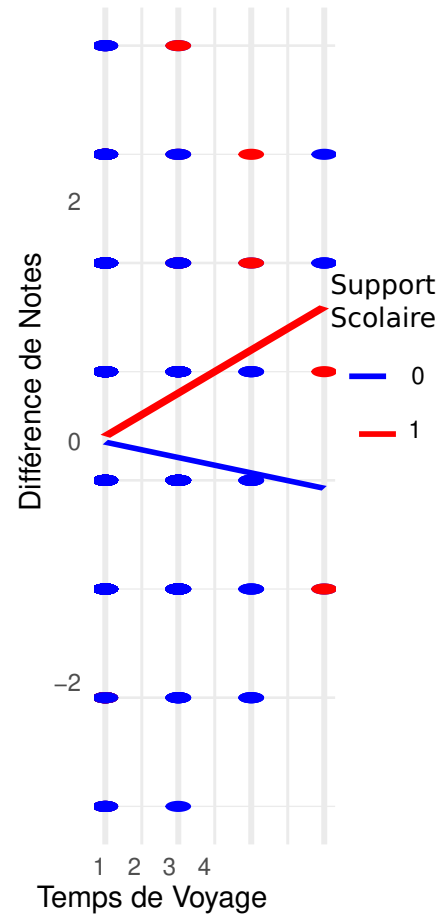
Interaction entre Temps de voyage et temps d'étude avec Différence de Notes



Interaction Temps de Voyage et Temps d'Étude

- Les interactions semblent suivre une logique cohérente : plus le temps passé en déplacement est important, moins l'effet bénéfique du temps d'étude est marqué.
- Cependant, pour les étudiants consacrant le plus de temps aux études, une augmentation du temps de voyage est associée à une amélioration plus notable des notes.

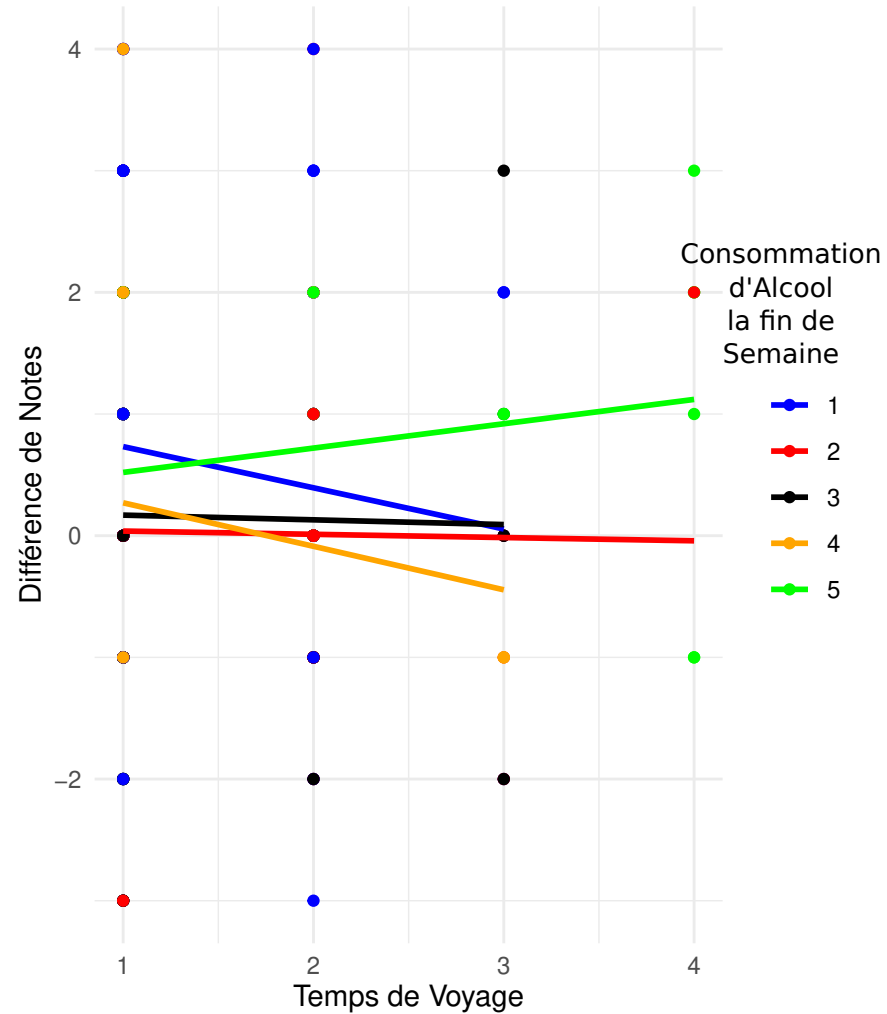
Interaction entre Temps de
Voyage et Support Scolaire
avec Différence de Notes



Interaction Temps de Voyage et Support Scolaire

- Le temps de voyage semble avoir un effet négatif sur les notes des étudiants, d'autant plus que le niveau du temps de voyage augmente
- Mais cela ne semble pas être le cas lorsqu'il y a du support scolaire.
- Pas d'explication pour ce résultat.

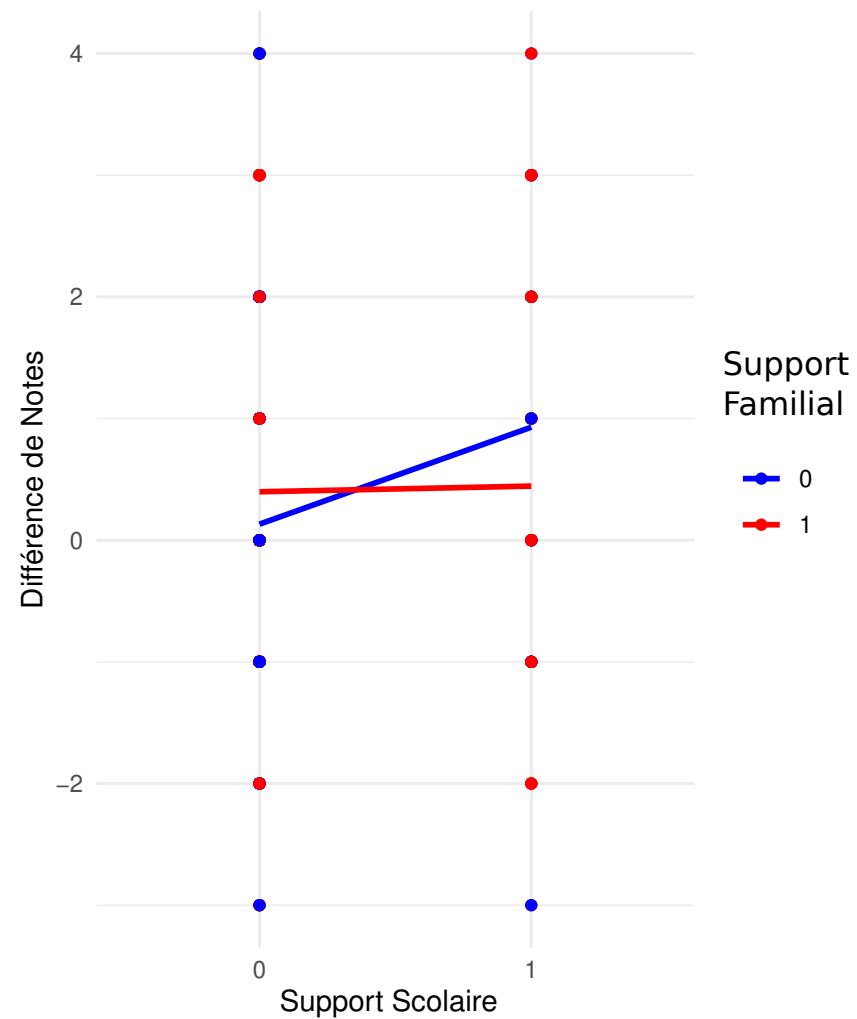
Interaction entre Temps de Voyage et Consommation d'Alcool la fin de semaine avec Différence de Notes



Interaction Temps de Voyage et Consommation d'alcool les fins de semaine

- La consommation d'alcool combiné au temps de voyage est généralement négativement proportionnel avec une bonne différence de note.
- Mais il y a exception, le niveau 4 du temps de voyage et le niveau 5 de la consommation
- Probablement du bruit, peu de données pour ces niveaux.

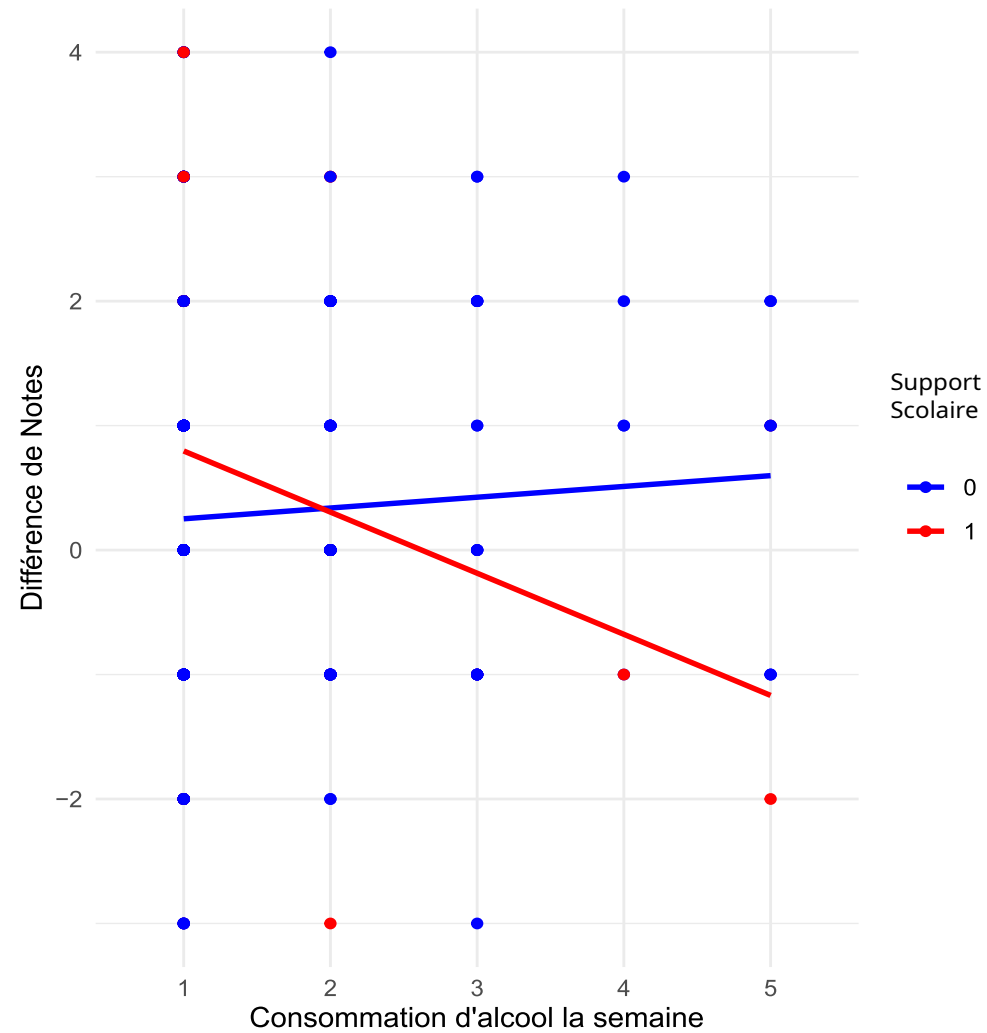
Interaction entre support scolaire et Support familial avec Différence de Notes



Interaction Support Scolaire et Support Familial

- Le support familial combiné au support scolaire n'a aucun effet sur la note.
- Le support scolaire seul a un effet positif.
- Probablement du bruit, peu de données pour ces niveaux.

Interaction entre consommation d'alcool la semaine et le support scolaire avec Différenc



Interaction Consommation d'alcool la semaine et Support Familial

- La consommation d'alcool la fin de semaine combiné au support scolaire est négative à partir du niveau 2 de consommation d'alcool.
- Si l'étudiant n'a pas de support scolaire, la consommation n'affecte pas vraiment la différence de note.
- Probablement du bruit, peu de données pour ces niveaux.

Troisième modèle

$$Y_{1d} = b_0 + \sum_{i=1}^8 b_i X_i + \sum_{i=1}^7 \sum_{j=i+1}^8 b_{ij} (X_i \cdot X_j) + \text{epsilon}$$

où

Y_{2m} = Note finale au cours de portugais ,

X_1 = temps d ' étude ,

X_2 = temps de voyage ,

X_3 = support scolaire ,

X_4 = support familial ,

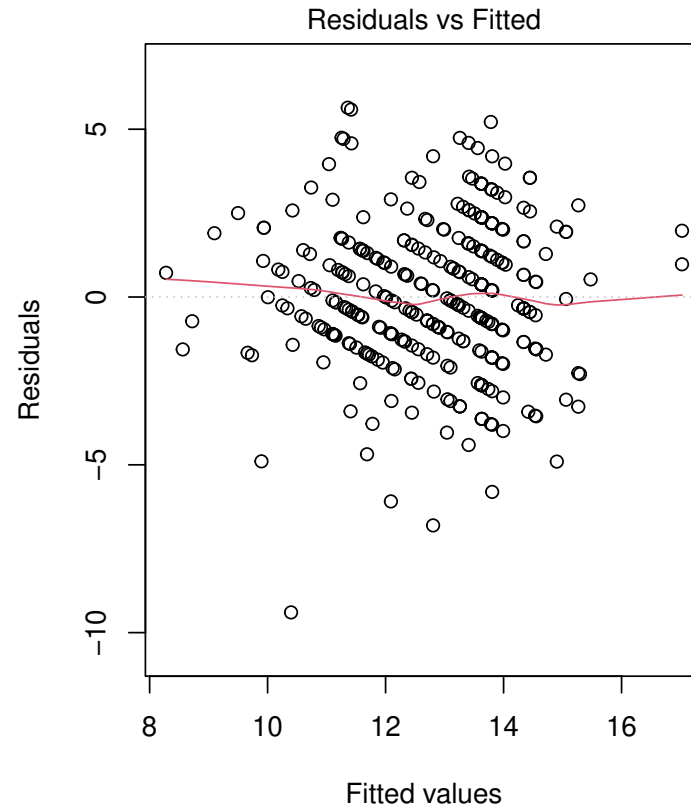
X_5 = cours privés ,

X_6 = internet ,

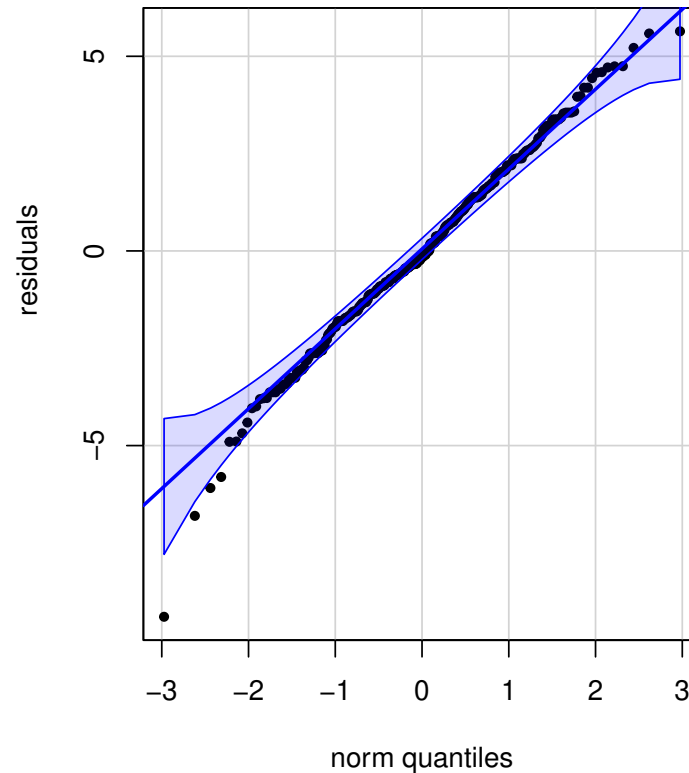
X_7 = Consommation d ' alcool les jours d ' école ,

X_8 = Consommation d ' alcool les fin de semaine

Homoscédasticité



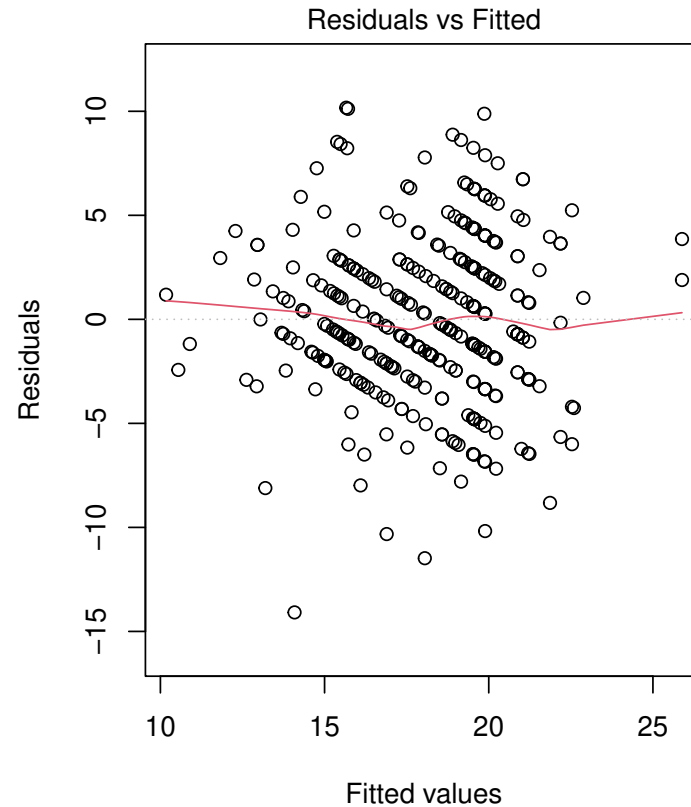
Normalité des résidus



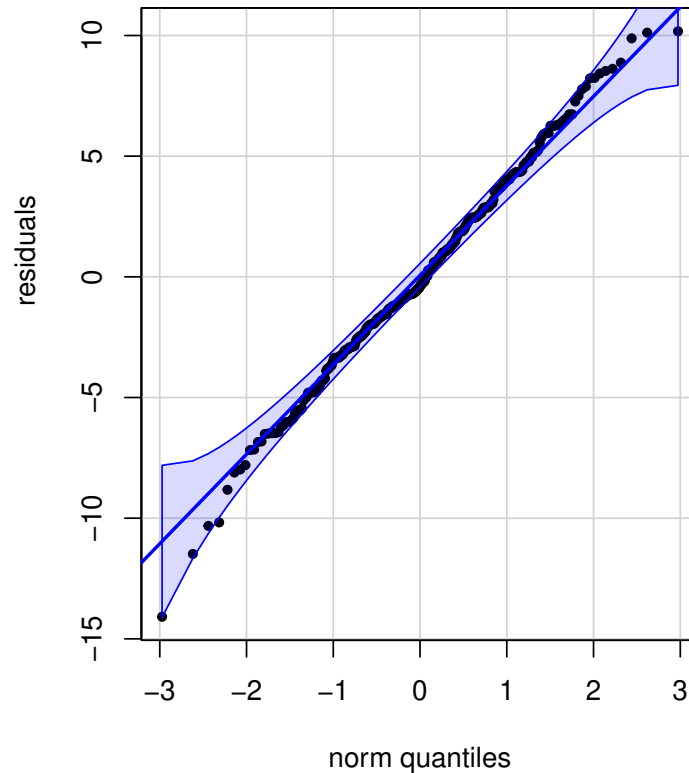
Normalité des résidus

- Test de Shapiro.
- $W = 0.99114$, avec une p-valeur de 0.03899.
- On fait une transformation de Box-Cox.
- On recrée le modèle.
- Et on reteste les 3 postulats.

Homoscédasticité



Normalité des résidus



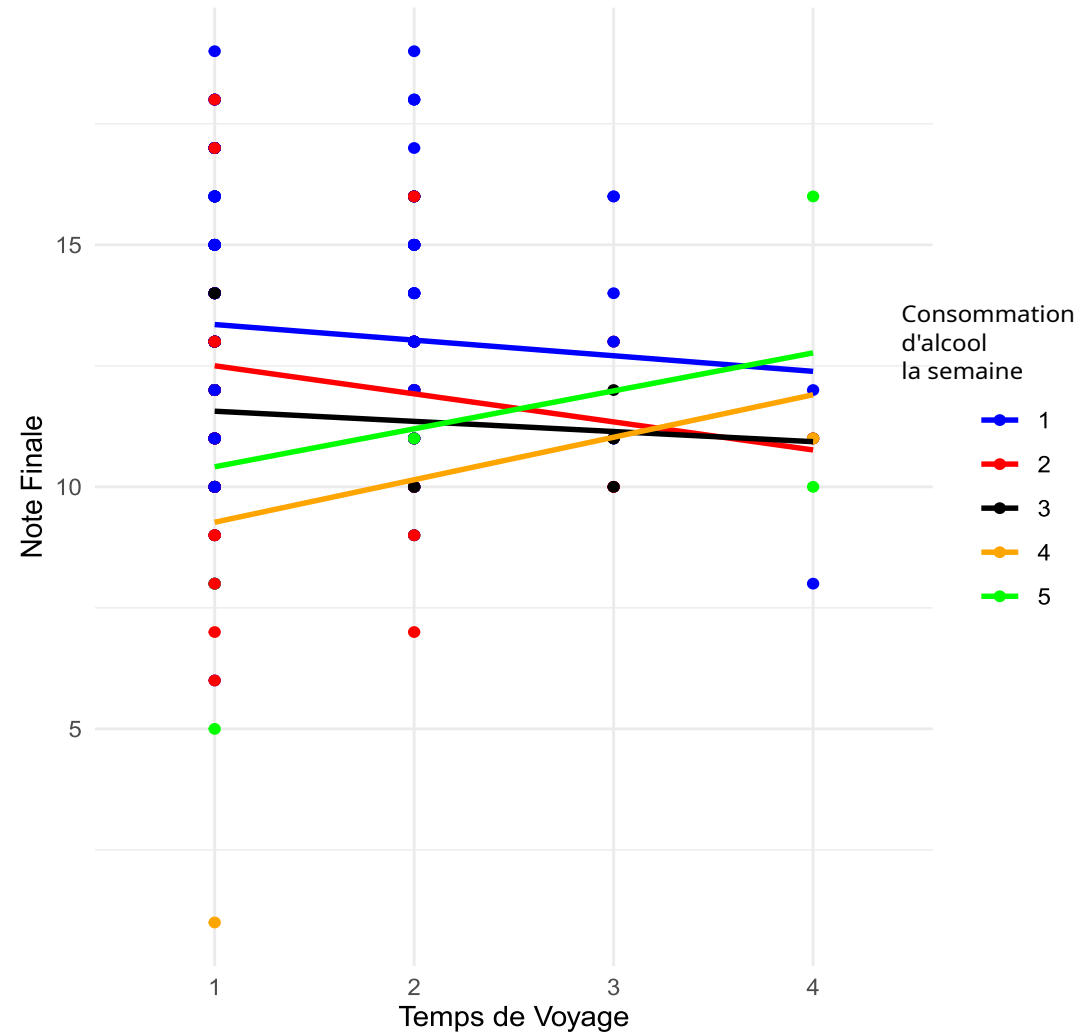
Normalité des résidus et égalité des variances

- Test de Shapiro.
- $W = 0.99515$, avec une p-valeur de 0.3667.
- Test de Durbin-Watson.
- $DW = 2.0743$, avec une p-valeur de 0.7465.

ANOVA

| Variable 1 | Variable 2 | P_Valeur |
|-------------------|-------------------------------|-----------------|
| Temps de voyage | Consommation d'alcool Semaine | 0.030593 |
| Support Familial | Internet | 0.007212 |

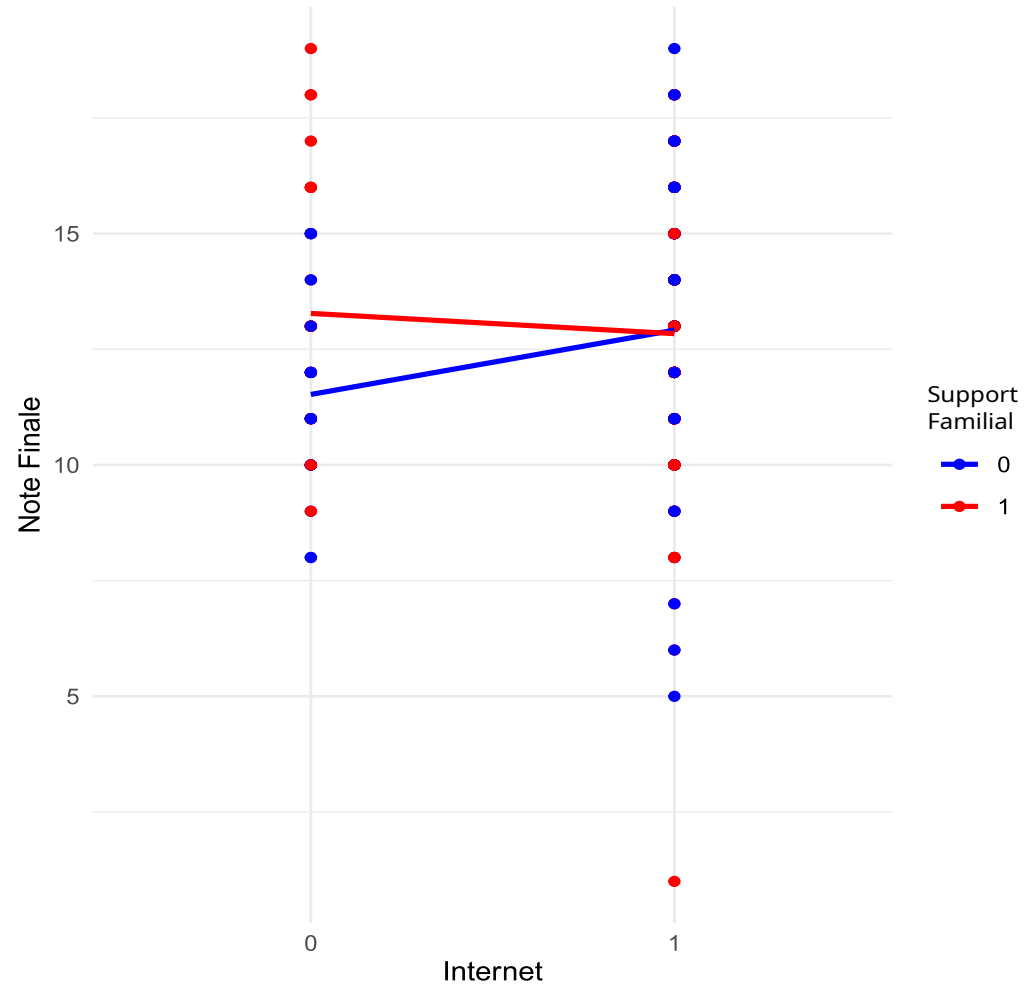
Interaction entre Temps de Voyage et Consommation d'Alcool la semaine avec Note Fin



Interaction Temps de Voyage et Consommation d'alcool la semaine

- Pour les niveau inférieur à 4 (pour la consommation d'alcool), le temps de voyage a un effet négatif et proportionnel.
- Par contre, il semble y avoir un effet positif à boire beaucoup d'alcool et avoir un long trajet.
- Probablement du bruit, peu de données pour ces niveaux.

Interaction entre la possession d'internet et le support familiale avec Note Finale



Interaction l'accès à Internet et Support Familial

- l'accès à Internet améliore la note uniquement en l'absence de support familial.
- En revanche, lorsque l'accès à Internet est combiné au support familial, cela nuit aux étudiants.

Quatrième modèle

$$Y_{1d} = b_0 + \sum_{i=1}^8 b_i X_i + \sum_{i=1}^7 \sum_{j=i+1}^8 b_{ij} (X_i \cdot X_j) + \text{epsilon}$$

où

Y_{2d} = Différence de note au cours de portugais ,

X_1 = temps d ' étude ,

X_2 = temps de voyage ,

X_3 = support scolaire ,

X_4 = support familial ,

X_5 = cours privés ,

X_6 = internet ,

X_7 = Consommation d ' alcool les jours d ' école ,

X_8 = Consommation d ' alcool les fin de semaine

Problème

- Le quatrième modèle ne respectait pas les 3 postulats
- Tentative avec Box-Cox, échoué.
- Tentative avec d'autres transformations, comme le sinus hyperbolique, échoué.
- Abandon de ce modèle

Conclusion

$$Y_{1m} = b_0 + b_1 X_1 + b_2 X_6 + b_3 X_8 + b_{12} (X_1 \cdot X_3) + b_{13} (X_1 \cdot X_4) + \text{epsilon}$$

$$Y_{1d} = b_0 + b_4 (X_2 \cdot X_3) + b_5 (X_2 \cdot X_4) + b_6 (X_2 \cdot X_8) + b_7 (X_3 \cdot X_4) + b_8 (X_3 \cdot X_8) + \text{epsilon}$$

$$Y_{2m} = b_0 + b_2 X_2 + b_1 X_1 + b_6 X_6 + b_7 X_7 + b_8 X_8 + b_9 (X_2 \cdot X_7) + b_{10} (X_4 \cdot X_6) + \text{epsilon}$$

où

Y_{1m} = Note Finale en mathématiques ,

Y_{1d} = Différence des notes en mathématique ,

Y_{2m} = Note Finale au cours de portugais ,

X_1 = temps d ' étude ,

X_2 = temps de voyage ,

X_3 = support scolaire ,

X_4 = support familial ,

X_5 = cours privés ,

X_6 = internet ,

X_7 = Consommation d ' alcool les jours d ' école ,

X_8 = Consommation d ' alcool les fins de semaine

Rappel de la première partie

Les prédicteurs significatifs sont :

- Temps d'étude.
- Internet.
- Consommation d'alcool les fins de semaine.
- Temps de voyage.
- Consommation d'alcool les journées d'écoles.

Les interactions significatives

- Temps d'étude avec le support scolaire ou le support familial.
- Mais cela pourrait être dû au fait que le temps d'étude est très significatif et positif.
- Alors que le support scolaire et support familial avait chacun un effet négatif.

Les interactions significatives

- Le temps de voyages avec le temps d'étude, ou le support scolaire, ou la consommation d'Alcool la fin de semaine.
- Le temps de voyages agit sur la différence de note en mathématique logiquement, à l'exception des niveau élevé de consommation d'alcool et temps d'étude. Et plus le temps de voyage augmente avec le support scolaire, plus la différence de notes est grande et positive.

Les interactions significatives

- Le support scolaire avec le support familial, ou la consommation d'alcool les journées d'école.
- La combinaison du support scolaire et familial n'a pas d'effet.
- S'il n'y a qu'un support scolaire, alors il y a un effet bénéfique sur la différence de notes en mathématique.
- La combinaison du support scolaire et la consommation à un effet proportionnellement négatif.
- Sans support scolaire, la consommation semble augmenter la note. Peut être est-ce du « bruit ».

Les interactions significatives

- Le temps de voyage avec la consommation d'alcool les jours de semaine.
- Relation proportionnellement négative, sauf pour les niveau 4 et 5 d'alcool. Cause possible? « Bruit ».
- La possession d'internet et le support familial.
- Il est marginalement plus avantageux de ne pas avoir internet et avoir du support familial.
- Si l'on a pas de support familial, il est avantageux d'avoir internet.

En Conclusion

- Les cours privés n'ont montré aucune incidence significative sur la performance des étudiants.
- Les trois modèles proposés offrent une vision globale de la performance étudiante et, par conséquent, une interaction qui s'avère significative dans un seul modèle pourrait être pertinente pour l'ensemble des performances.
- Cependant, les résultats pourraient être influencés ou même erronés, en raison de la nature observationnelle de l'étude et de l'absence de données pour l'année précédente.