جامعة مولاي إسماعيل
UNIVERSITÉ MOULAY ISMAÏL
كلية العلوم والتقنيات
FACULTÉ DES SCIENCES ET TECHNIQUES

Département : d'Informatique
Master : Systèmes d'Information Décisionnels et Imagerie
Module : Data-minig
Année universitaire : 2022/2023
Compte rendu de mini projet sous le thème :

Mini-projets Data-Ming et Machine Learning dans l'éducation

Prédiction de résultat des élèves en se basant sur des caractères sociaux

Réaliser par :

BOUARAFA Badr

Encadrer par :

Professeur : Mohamed Sabiri

Table des matières

Présentation de la base de données :

Source :

Kaggle https://www.kaggle.com/uciml/student-alcohol-consumption

Présentation des données :

Les données ont été obtenues dans le cadre d'une enquête auprès des élèves des cours de mathématiques et de langue portugaise du secondaire. Il contient de nombreuses informations intéressantes sur les élèves sur les plans social et démographiques, du genre et leurs études.
Le jeu de données est composé de deux fichiers Excel, student-mat.csv (cours de mathématiques) et student-por.csv (cours de langue portugaise).
Les variables des ensembles de données sont :

- school - école de l'élève (binaire : 'GP' - Gabriel Pereira ou 'MS' - Mousinho da Silveira)
- sex - sexe de l'élève (binaire : 'F' - femme ou 'M' - homme)
- age - âge de l'élève (numérique : de 15 à 22)
- addresse - type d'adresse du domicile de l'élève (binaire : 'U' - urbain ou 'R' - rural)
- famsize - taille de la famille (binaire : 'LE3' - inférieur ou égal à 3 ou 'GT3' - supérieur à 3)
- Pstatus - statut de cohabitation des parents (binaire : 'T'- vivant ensemble ou 'A' - séparé)
- Medu - éducation de la mère (numérique : 0 - aucun, 1 - enseignement primaire (4e année), 2 - 5e à 9e année, 3 - enseignement secondaire ou 4 - enseignement supérieur)
- Fedu - éducation du père (numérique : 0 - aucun, 1 - enseignement primaire (4e année), 2 - 5e à 9e année, 3 - enseignement secondaire ou 4 - enseignement supérieur)

× Mjob - travail de la mère (nominal : 'teacher', 'health', 'services' (par exemple administratifs ou policiers), 'at_home' ou 'other')
× Fjob - travail du père (nominal : 'teacher', 'health', 'services' (par exemple administratifs ou policiers), 'at_home' ou 'other')
× reason - raison de choisir cette école (nominale : 'home' =proche de domicile , 'reputation '=réputation de l'école, 'cours'= préférence de cours ou 'autre')
× guardian- tuteur de l'élève (nominal : 'mother ', 'father' ou ' autre ')
× traveltime- temps de trajet domicile-école (numérique : 1 si <15 min, 2 si 15 à 30 min, 3 si 30 min à 1 heure ou 4 si 1 heure)
× studytime - temps d'étude hebdomadaire (numérique : 1 si <2 heures, 2 si 2 à 5 heures, 3 si 5 à 10 heures ou 4 si >10 heures)
× failures - nombre d'échecs de classe passés (numérique : n si 1<=n<3, sinon 4)
× schoolsup - soutien pédagogique supplémentaire (binaire : 'yes' ou 'no')
× famsup - soutien éducatif familial (binaire : 'yes' ou 'no')
× paid - cours payants supplémentaires dans la matière du cours (mathématiques ou portugais) (binaire : 'yes' ou 'no')
× activities - activités extra-scolaires (binaire : 'yes' ou 'no')
× nursery - a fréquenté une école maternelle (binaire : 'yes' ou 'no')
× higher - souhaite faire des études supérieures (binaire : 'yes' ou 'no')
× internet - Accès Internet à la maison (binaire : 'yes' ou 'no')
× romantic - avec une relation amoureuse (binaire : 'yes' ou 'no')
× famrel - qualité des relations familiales (numérique : de 1 - très mauvaise à 5 - excellente)
× freetime- temps libre après l'école (numérique : de 1 - très faible à 5 - très élevé)
× goout - sortir avec des amis (numérique : de 1 - très faible à 5 - très élevé)
× Dalc - consommation d'alcool au travail (numérique : de 1 - très faible à 5 - très élevée)
× Walc - consommation d'alcool le week-end (numérique : de 1 - très faible à 5 - très élevée)
× Health - état de santé actuel (numérique : de 1 - très mauvais à 5 - très bon)
× absences - nombre d'absences scolaires (numérique : de 0 à 93)

Ces notes sont liées à la matière du cours, mathématiques ou portugais :

× G1 - note de première période (numérique : de 0 à 20)
× G2 - note de deuxième période (numérique : de 0 à 20)
× G3 - note de troisième période (numérique : de 0 à 20)

Variable	Valeur
school	GP,MS

sex	F,M
address	R,U
famsize	GT3,LE3
Pstatus	A,T
Mjob Fjob	at_home,health,other,services,teacher
reason	course,home,other,reputation
guardian	father,mother,other
schoolsup famsup paid activities nursery higher internet romantic	no,yes

Objectif :

On va utiliser ce jeu de données "student-mat.csv" (cours de mathématiques) pour certains EDA (Exploration de Données) et essayer de prédire si un élève réussira ou non en se basant sur des critères sociaux.
J'ai calculer la moyenne de 'G1','G2' et 'G3' dans une nouvelle colonne 'Avirage'et j'ai classé ces élèves en deux catégories, 'Fail' et 'pass', en fonction de leurs résultats final.
Ensuite, j'ai explorer et analysé quelques caractéristiques qui ont une influence significative sur la performance finale des élèves. Enfin, j'ai créé divers modèles d'apprentissage automatique pour prédire la classe des élèves et j'ai comparé les performances des modèles.

Outil :

Rattle de Togawar et R

Instalation :

1/ Aller au site : https://cran.r-project.org/
Télécharger et installer le programme: R 3.6.2 (December, 2019)
Télécharger et installer le programme: Rtools35.exe
2/ Aller au site: https://www.msys2.org/
Télécharger et installer le programme: msys2-x86_64-20231026.exe
Ouvrez le terminal MSYS2 et installez les dépendances nécessaires avec la commande suivante :
pacman -S mingw-w64-x86_64-gtk2 mingw-w64-x86_64-pango mingw-w64-x86_64-cairo
Cette commande installe les bibliothèques GTK+, Pango et Cairo nécessaires pour RGtk2.
3/ Configuration les variables d'environnement:

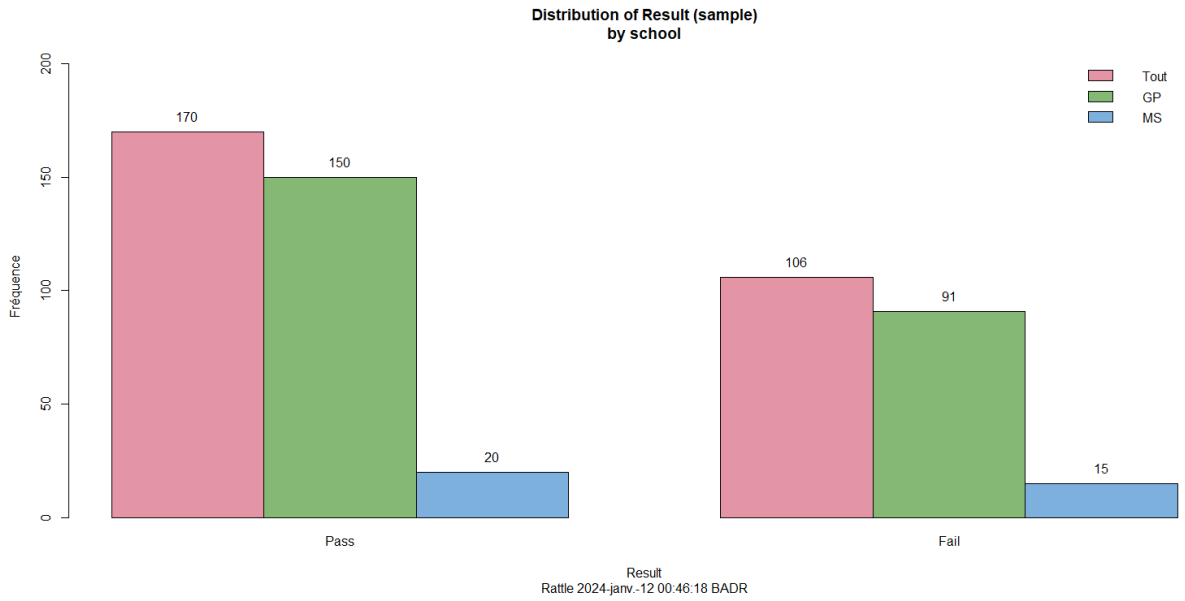
Soit manuellement: Windows ->Système -> Paramètres avancés de système -> Variable d'environnement -> Path -> modifier
Ajouter les liens:
C:\Rtools
C:\Rtools\bin
C:\Rtools\lib
C:\Rtools\include
C:\Rtools\mingw_64
C:\Rtools\mingw_64\bin
C:\Rtools\mingw_64\lib
C:\Rtools\mingw_64\include
C:\msys64
C:\msys64\mingw64
C:\msys64\mingw64\bin
C:\msys64\mingw64\lib
C:\msys64\mingw64\include
4/ Sous R Utiliser les commandes:
>install.packages("rattle")
> install.packages("https://cran.r-project.org/src/contrib/Archive/RGtk2/RGtk2_2.20.36.tar.gz", repos = NULL)
>install.packages("Rcpp")

Préparation des données

*Creation de deux colonnes "Average"
Et "Result"
*Calculer la moyenne de "G1" , "G2" et "G3" dans la colonne "Average"
*Atribuéer à chaque élève "Fail" si sa moyenne<0 ou "Pass" si sa moyenne >0
Code R:

```r
#Vous pouvez définir le répertoire de travail
setwd("C:/Users/BADR/Desktop/Nouveau dossier/img")

# Charger la bibliothèque 'dplyr' pour la manipulation des données
library(dplyr)

# Lire la base de données depuis le fichier CSV
data <- read.csv("C:/Users/BADR/Documents/R/student-mat.csv")

# Remplacer les colonnes 'G1', 'G2' et 'G3' par leur moyenne dans une nouvelle colonne 'Average'
data$Average <- rowMeans(data[, c('G1', 'G2', 'G3')], na.rm = TRUE)

# Créer une nouvelle colonne 'Result' avec les valeurs 'Fail' si 'Average' < 10 et 'Pass' sinon
data$Result <- ifelse(data$Average < 10, 'Fail', 'Pass')

# Afficher les premières lignes de la base de données résultante
head(data)

# Enregistrer la base sous le nom "student-mat-prepared.csv"
write.csv(data, file = "student-mat-prepared.csv", row.names = FALSE)
```

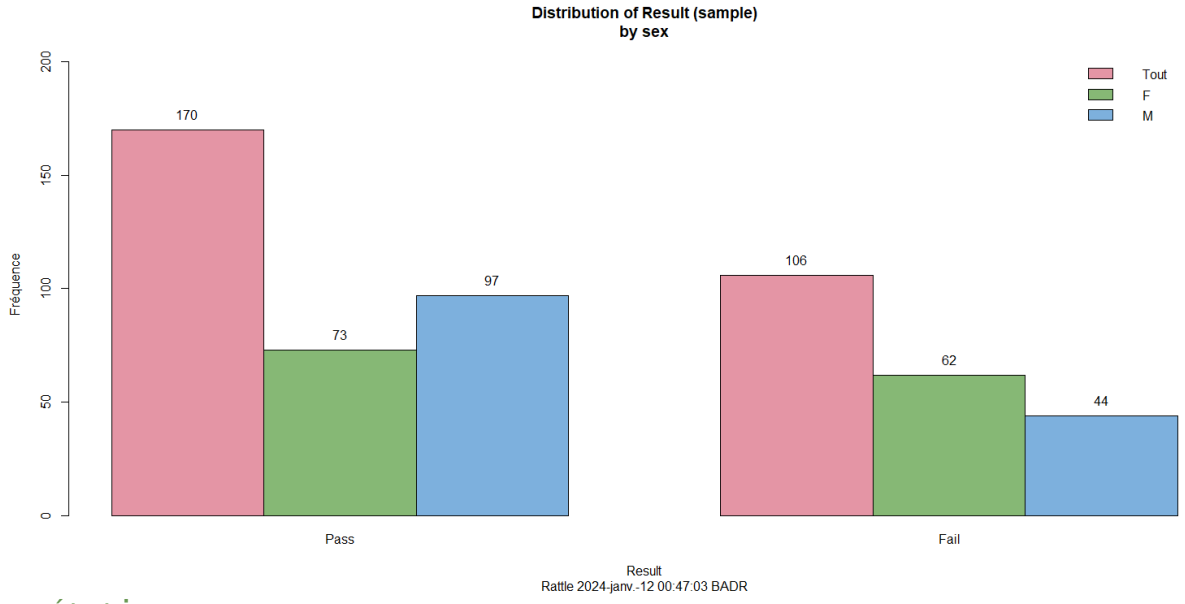
Explorations des données :



Interprétation

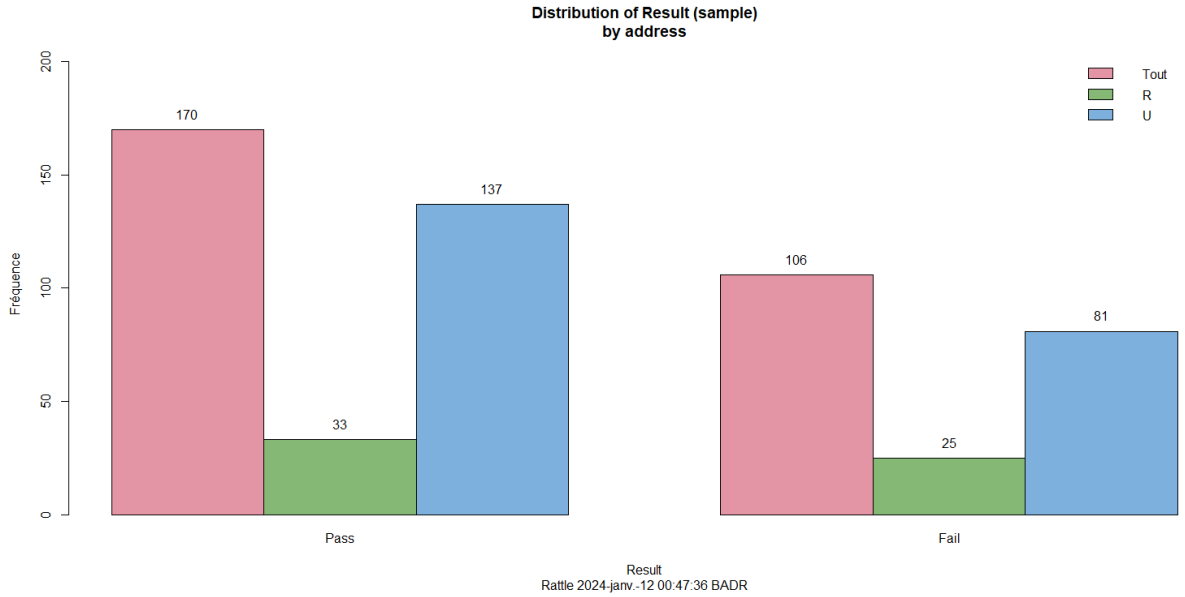
On peut voir que l'école GP a un nombre total d'élèves supérieur à l'école MS. En outre, l'école GP a un taux de réussite plus élevé que l'école MS.
Il est possible que:

- L'école GP est une école plus ancienne que l'école MS.
- L'école GP a un corps professoral plus expérimenté que l'école MS.



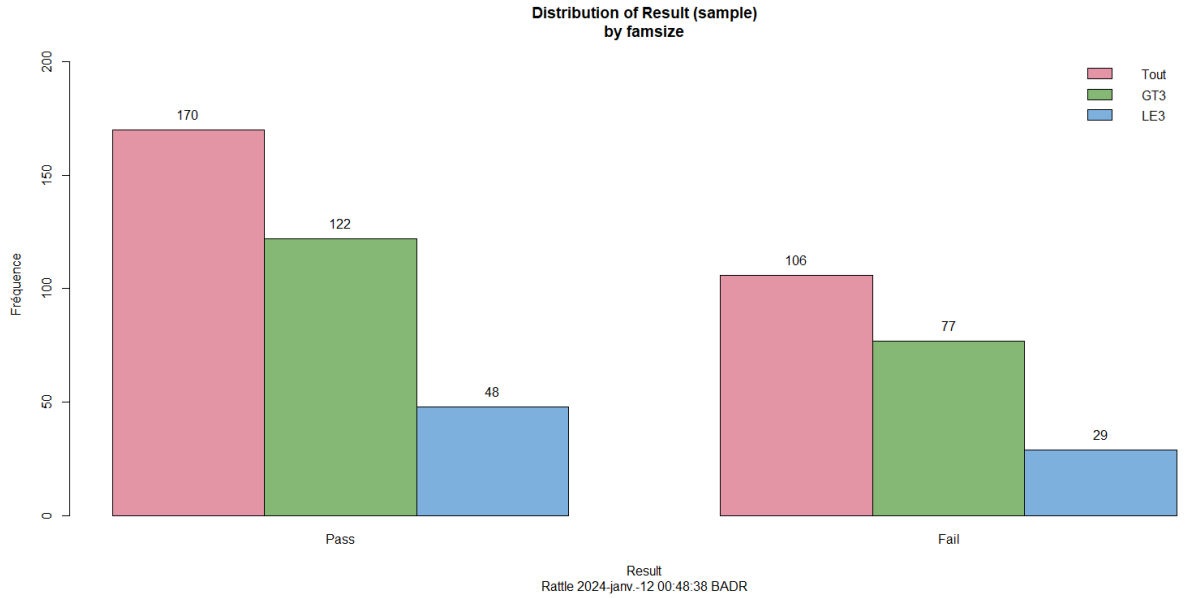
Interprétation

On peut voir que les filles ont un taux de réussite plus élevé que les garçons. La proportion de filles qui réussissent est de 73 %, tandis que la proportion de garçons qui réussissent est de 62 %.
Possible que:
Les filles sont plus motivées que les garçons à réussir dans cet examen



Interprétation

La distribution des résultats par adresse montre que le taux de réussite est plus élevé dans les zones urbain que dans les zonne rural.

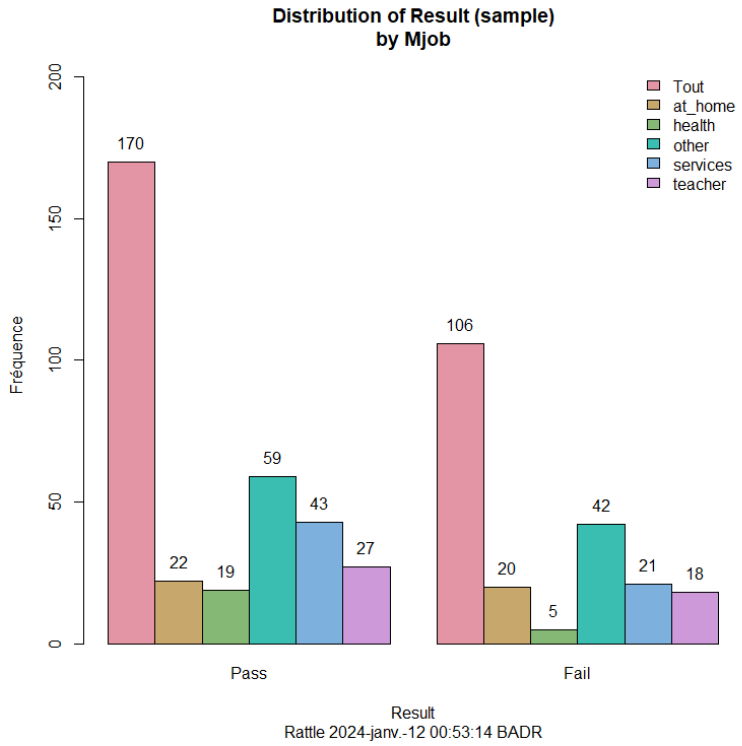


Interprétation

les familles de plus de 3 personnes ont plus de chances de réussir le test que les familles de moins de 3 personnes. En effet, les familles de plus de 3 personnes ont généralement plus de ressources et de soutien, ce qui peut leur donner un avantage dans le test.

Cependant, il est important de noter que cette image ne montre qu'une tendance générale. Il existe des familles de moins de 3 personnes qui réussissent le test, et des familles de plus de 3 personnes qui échouent.

Il est possible que:
Famille nombreuse = plus de chances de réussite, mais pas toujours

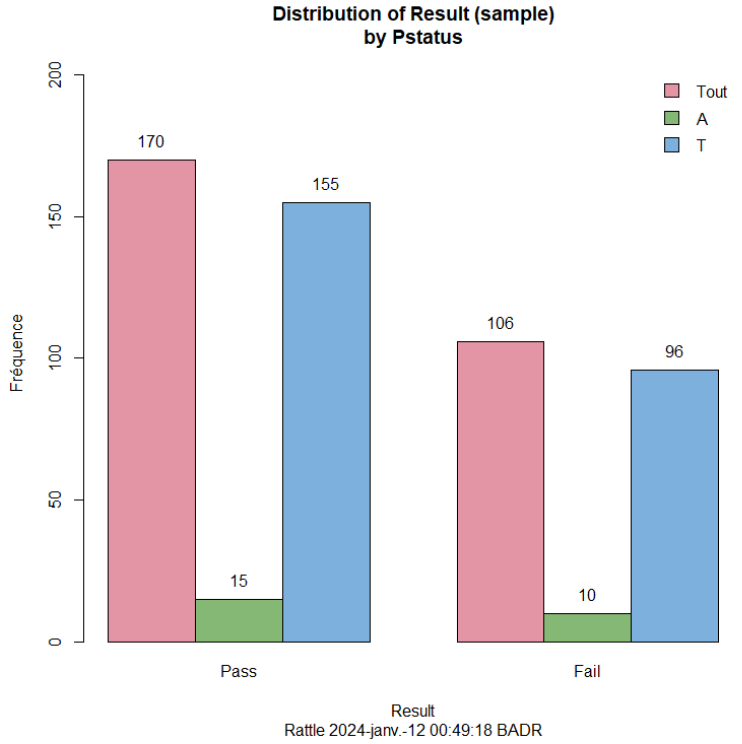


Interprétation

Les élèves dont la mère est enseignante ont tendance à avoir de meilleurs résultats que les élèves dont la mère a un autre métier.

Conclustion :

e métier d'enseignante est un facteur qui peut contribuer à la réussite scolaire des enfants, mais il n'est pas le seul facteur. D'autres facteurs, tels que les capacités cognitives des enfants, leurs motivations et leurs conditions de vie, peuvent également influencer les résultats scolaires.

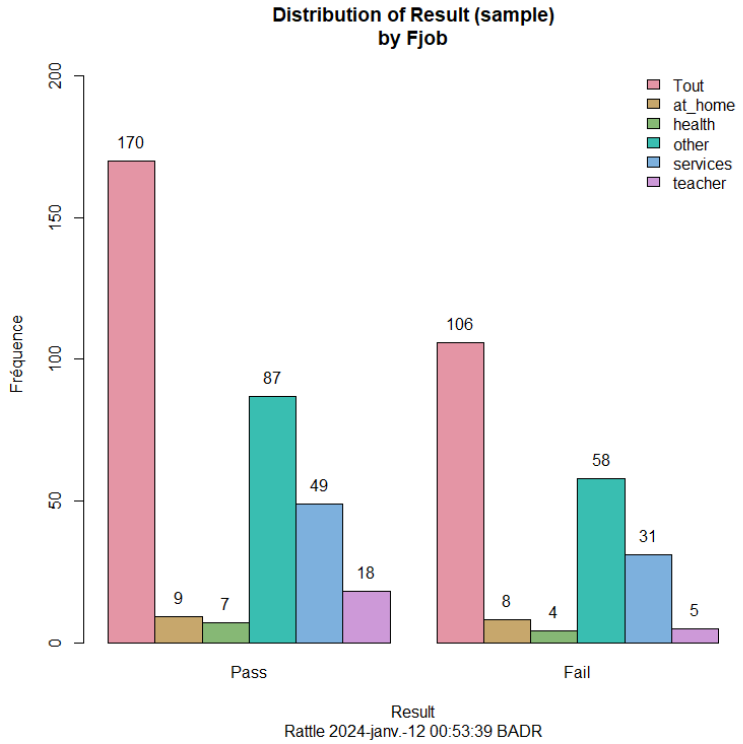


Interprétation

Le graphique montre que les enfants de familles nombreuses ont un taux de réussite légèrement plus élevé que les enfants de familles monoparentales. La différence est d'environ 5 %. Cela signifie que, en moyenne, un enfant de famille nombreuse a 5 % de chances de plus de réussir le test qu'un enfant de famille monoparentale.

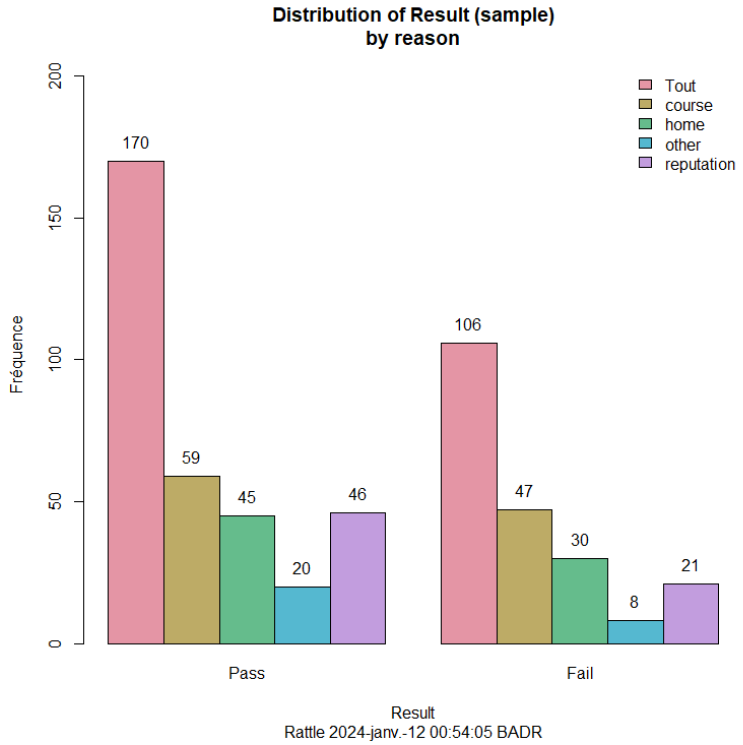
Il est possible que:

Les résultats du test sont meilleurs pour les enfants de familles nombreuses, mais la différence est minime.



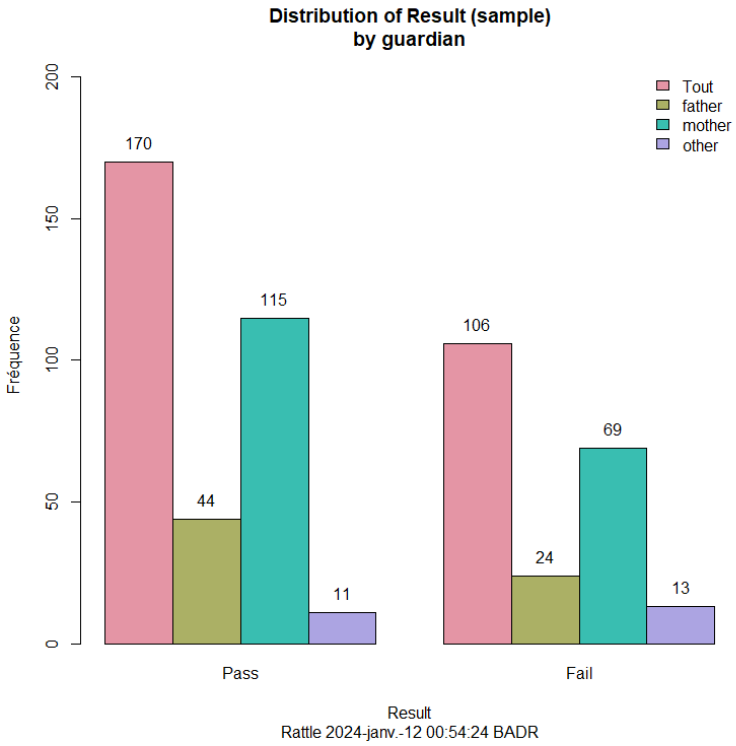
Interprétation

les enfants qui ont un père présent ont un taux de réussite légèrement plus élevé que les enfants qui n'ont pas de père présent. La différence est d'environ 5 %. Cela signifie que, en moyenne, un enfant qui a un père présent a 5 % de chances de plus de réussir le test qu'un enfant qui n'a pas de père présent.
possible :
La présence d'un père est un facteur important de réussite au test, mais il n'est pas le seul facteur



Interprétation

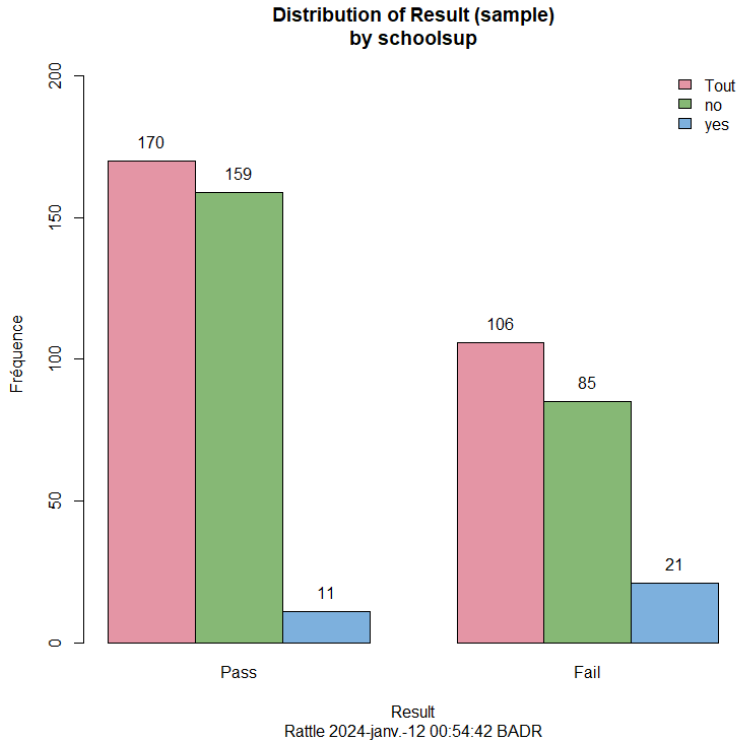
Les élève qui ont choisi l'école pour les raisons (proche de domicile , réputation de l'école, préférence de cours ou autre) ont presque la même chance de réussite



Interprétation

la mère est le tuteur de 170 élèves, suivi du père avec 115 élèves. Les autres tuteurs comprennent les grands-parents, les tuteurs et les autres membres de la famille

il est important de noter que cette information ne doit pas être interprétée comme une preuve que les mères sont intrinsèquement meilleures que les pères



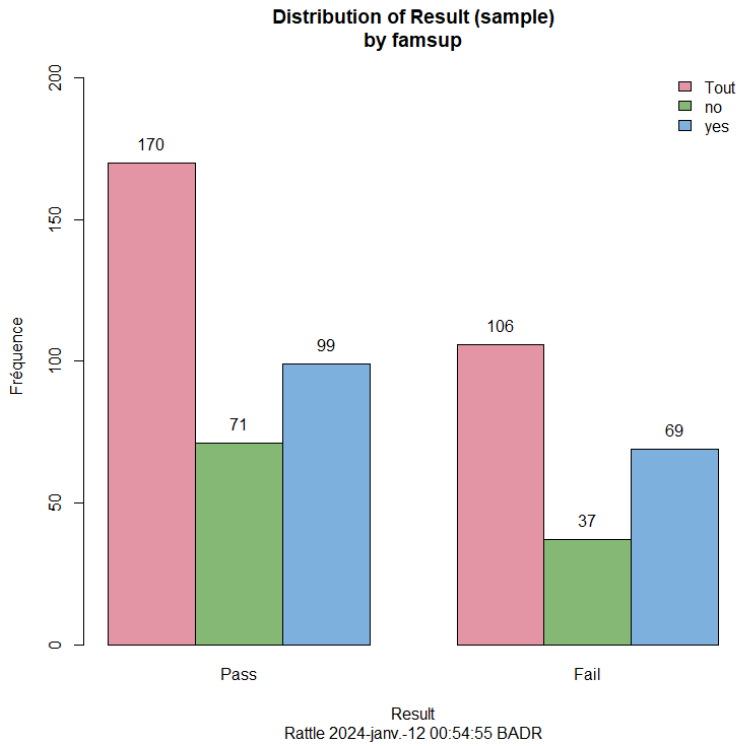
Interprétation

Les élèves qui ont reçu un soutien pédagogique supplémentaire ont plus de chances de réussir au test.

les élèves qui ont reçu un soutien pédagogique supplémentaire ont un taux de réussite de 85 %, contre 66 % pour les élèves qui n'ont pas reçu de soutien. Cela signifie que, en moyenne, un élève qui a reçu un soutien pédagogique supplémentaire a 29 % de chances de plus de réussir au test qu'un élève qui n'en a pas reçu.

Conclusion :

Le soutien pédagogique supplémentaire augmente les chances de réussite au test

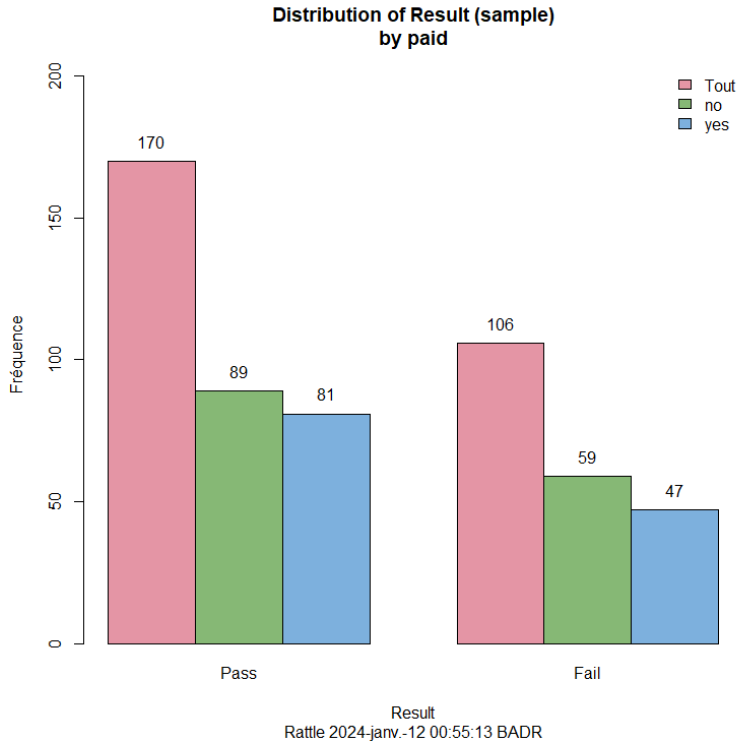


Interprétation

Les élèves qui ont reçu un soutien éducatif familial ont plus de chances de réussir au test, mais la différence est minime

Possible:

Le soutien éducatif familial peut augmenter les chances de réussite au test, mais ce n'est pas un facteur décisif

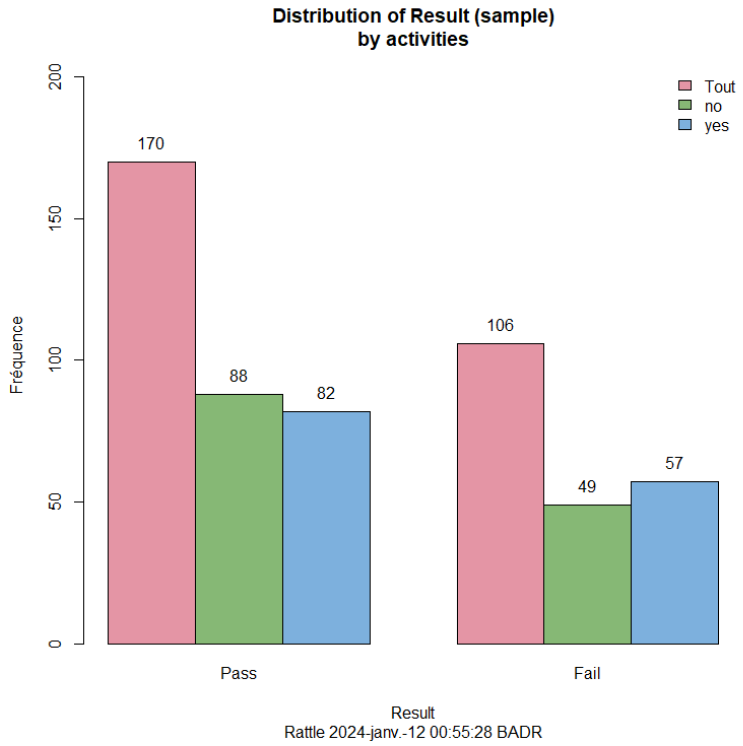


Interprétation

les élèves qui ont reçu un soutien éducatif familial ont un taux de réussite de 71 %, contre 69 % pour les élèves qui n'ont pas reçu de soutien. Cela signifie que, en moyenne, un élève qui a reçu un soutien éducatif familial a 2 % de chances de plus de réussir au test qu'un élève qui n'en a pas reçu.

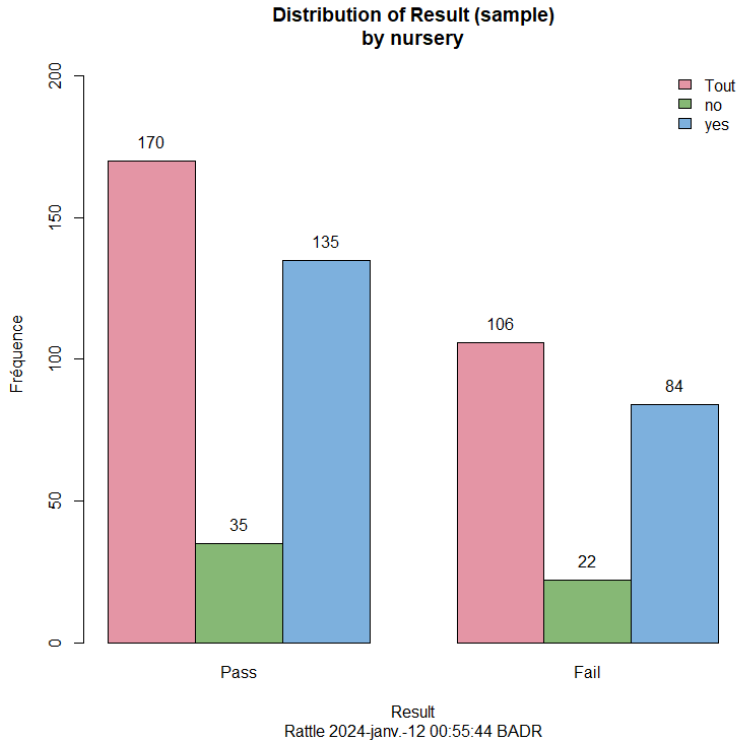
Possible:

Le soutien éducatif familial peut augmenter les chances de réussite au test, mais ce n'est pas un facteur décisif.



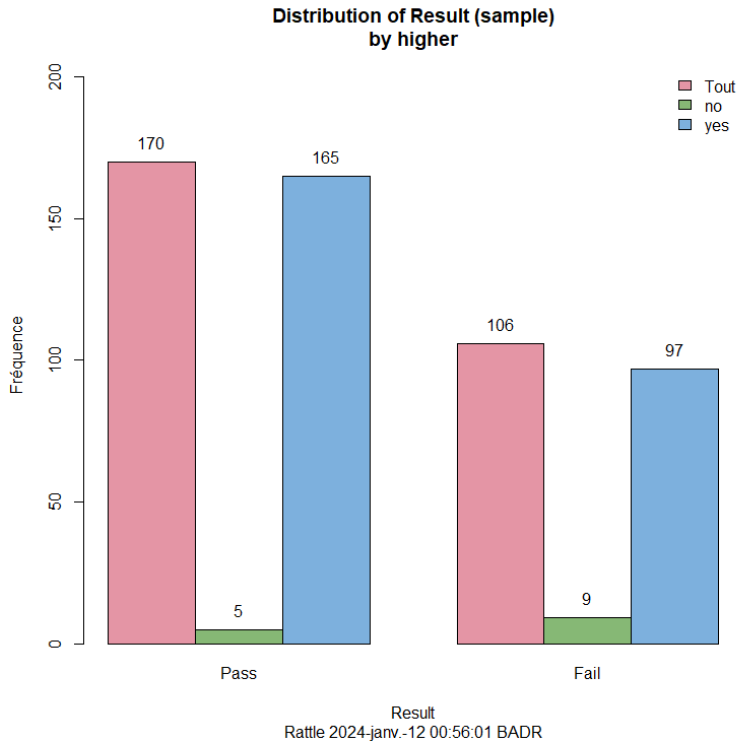
Interprétation

Les élèves qui participent à des activités extra-scolaires ont tendance à avoir de meilleurs résultats scolaires, en particulier les élèves qui ont des difficultés.
Conclusion :La participation à des activités extra-scolaires est un facteur qui peut contribuer à la réussite scolaire des enfants.



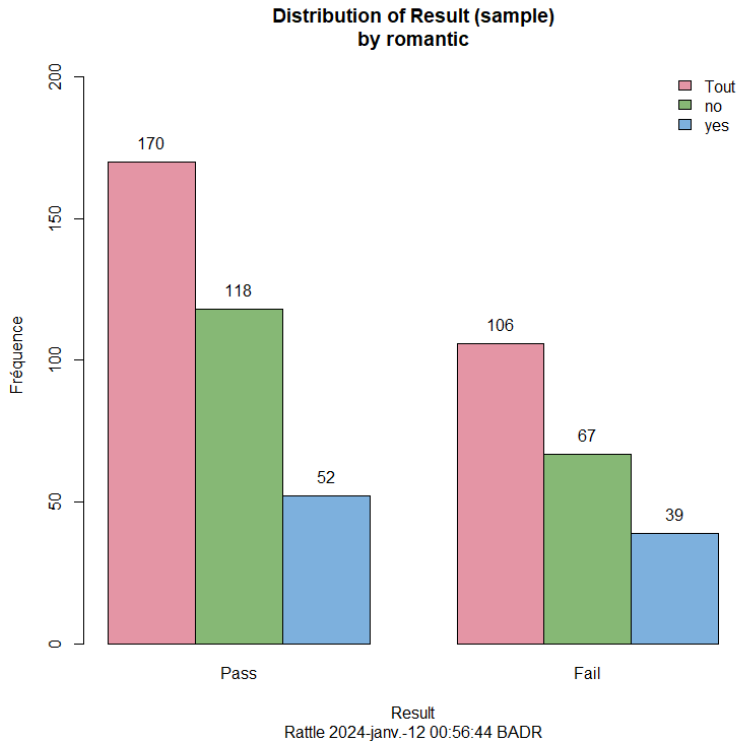
Interprétation

Les élèves qui ont fréquenté une école maternelle ont plus de chances de réussir au test, mais la différence n'est pas énorme.
Cette interprétation est plus nuancée. Elle indique que la différence entre les taux de réussite des deux groupes d'élèves est significative, mais qu'elle n'est pas énorme. Il est possible que d'autres facteurs, tels que le soutien familial ou l'environnement socio-économique, soient également importants pour la réussite au test



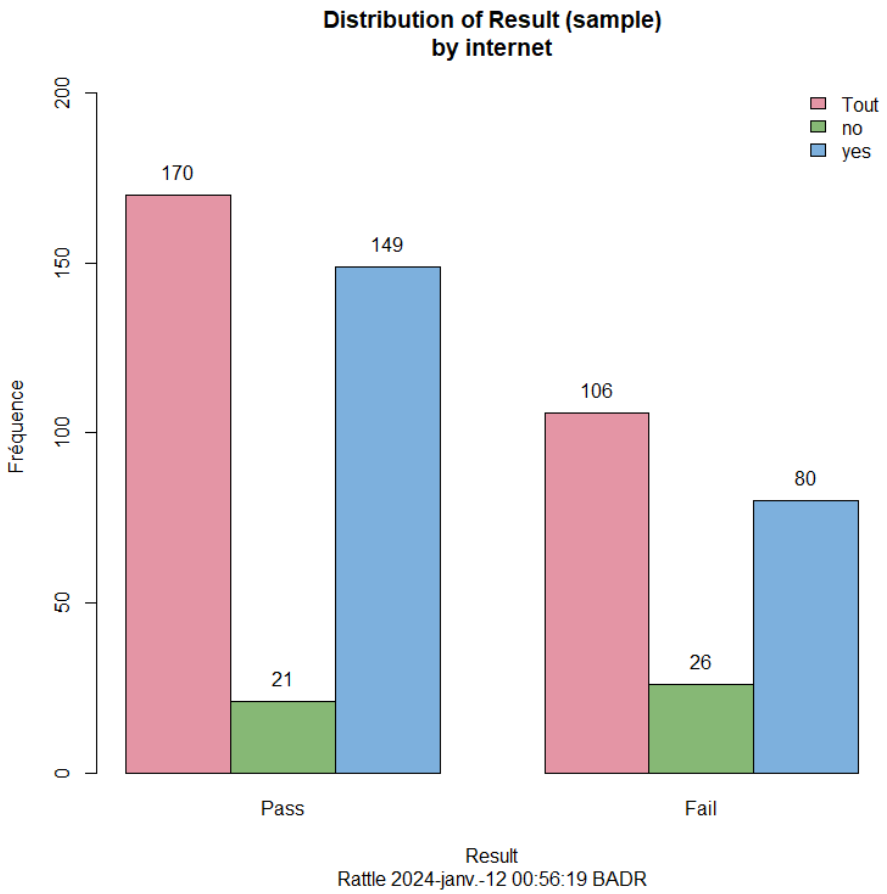
Interprétation

Les élèves qui souhaitent faire des études supérieures sont plus susceptibles de réussir au test, mais la différence n'est pas énorme.



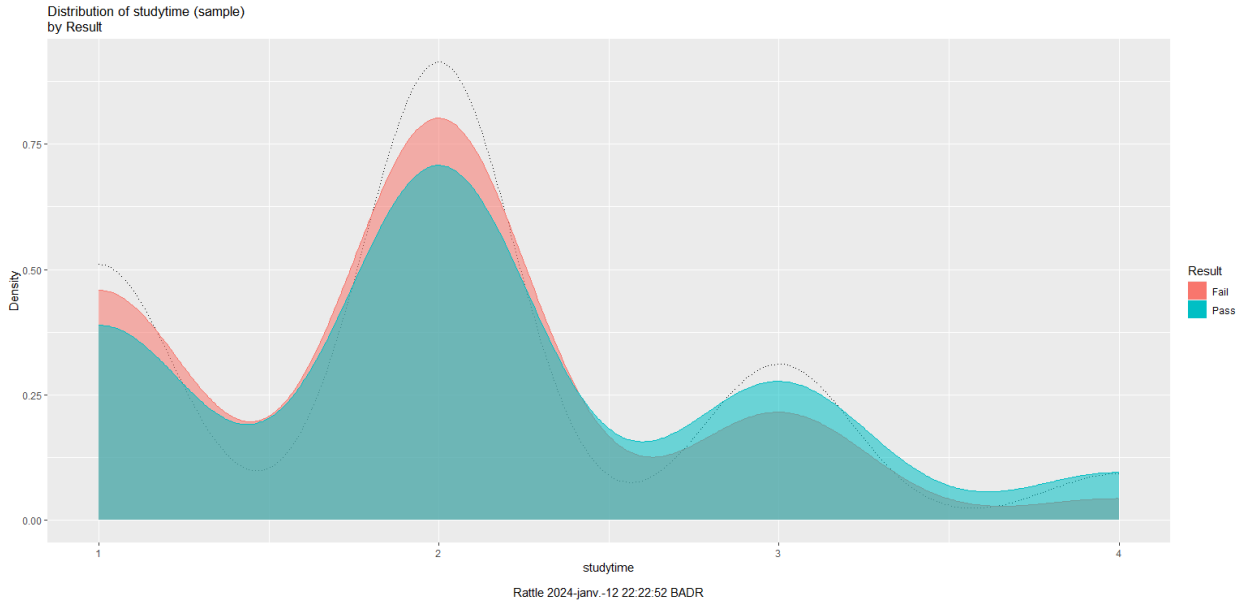
Interprétation

Les élèves qui ne sont pas dans une relation amoureuse sont plus susceptibles de réussir au test, mais la différence n'est pas énorme.



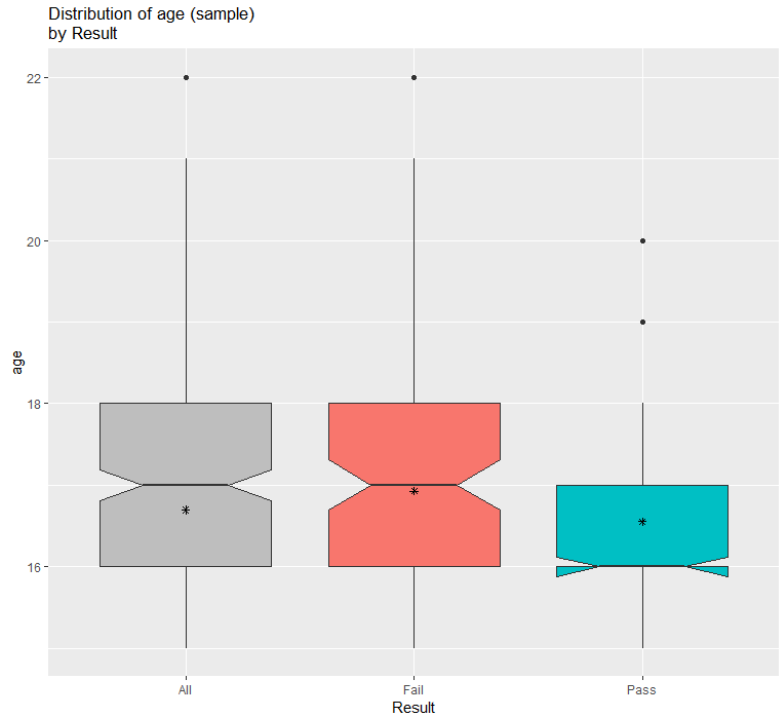
Interprétation

La majorité des élèves utilisent Internet pour les études mais le taux de réussite est presque le même pour les deux classes



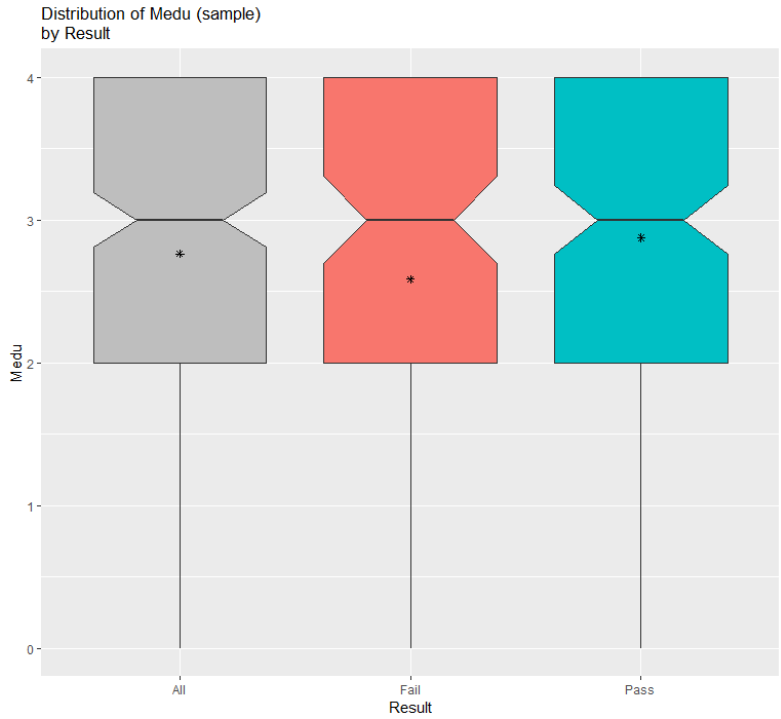
Interprétation

Les élèves qui ont échoué à l'examen ont tendance à avoir étudié moins que les élèves qui ont réussi



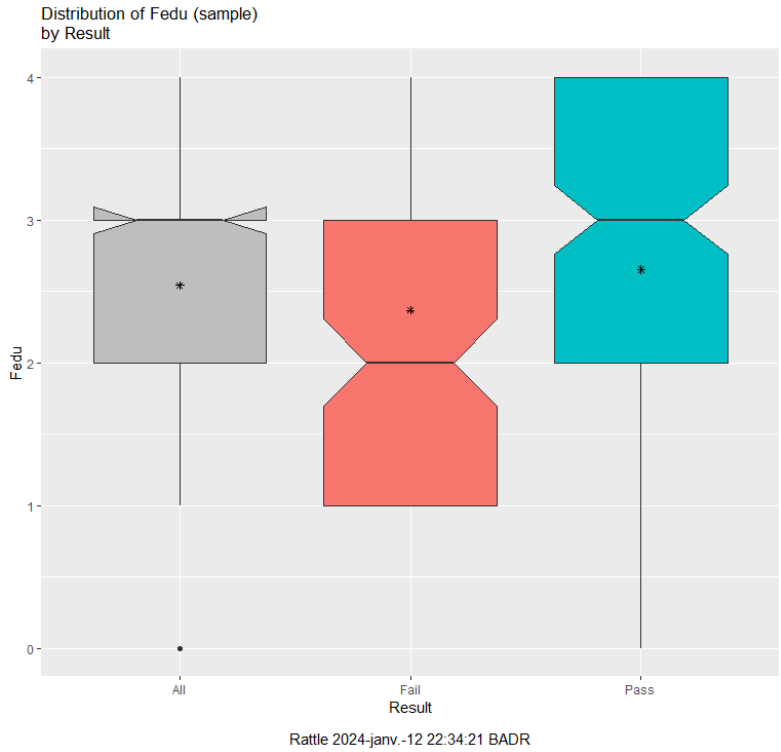
Interprétation

La moitié des élèves qui ont échoué avaient moins de 18 ans, contre seulement 20 % des élèves qui ont réussi.
La majorité des élèves qui ont réussi avaient entre 18 et 20 ans.
Une petite minorité d'élèves qui ont réussi avaient plus de 20 ans.
possible:
'il existe une corrélation entre l'âge et les résultats aux examens. Cependant, il est important de noter que cette corrélation n'est pas causale



Interprétation

Les élèves dont les mères ont un niveau d'éducation plus élevé ont tendance à avoir de meilleurs résultats aux examens.
Conclusion : Le niveau d'éducation de la mère est un facteur important pour la réussite scolaire des enfants.

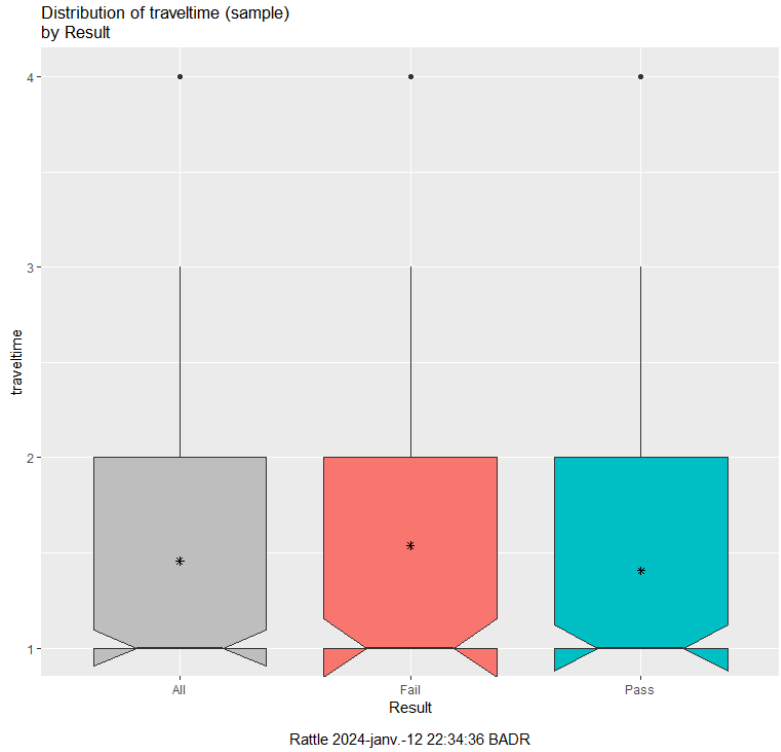


Interprétation

Les élèves qui ont reçu une éducation préscolaire ont plus de chances de réussir aux examens que les élèves qui n'en ont pas reçu.

Conclusion :

L'éducation préscolaire est un facteur important pour la réussite scolaire des enfants.

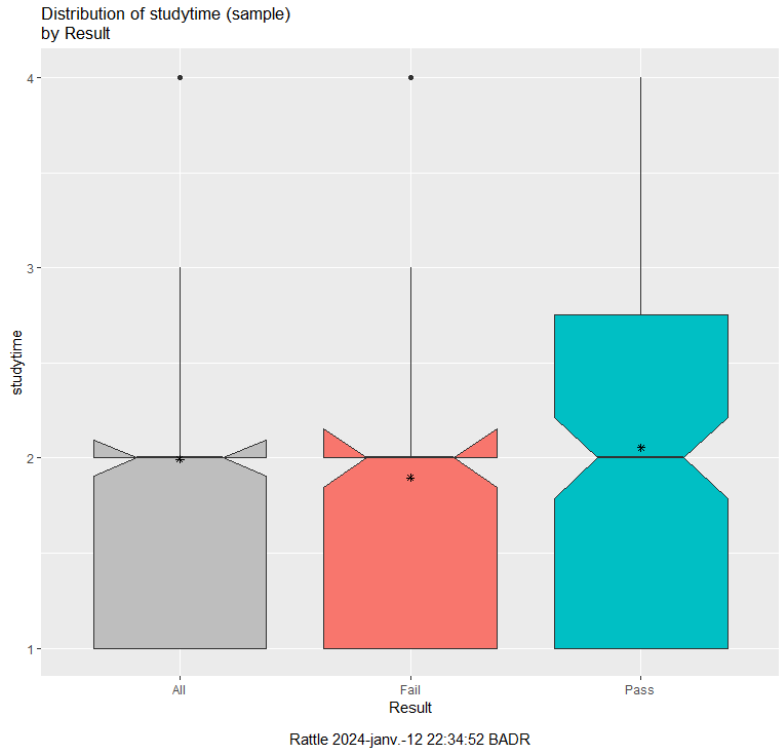


Interprétation

Les élèves qui ont un temps de trajet domicile-école plus court ont tendance à avoir de meilleurs résultats aux examens.

Conclusion :

Le temps de trajet domicile-école est un facteur important pour la réussite scolaire des enfants.

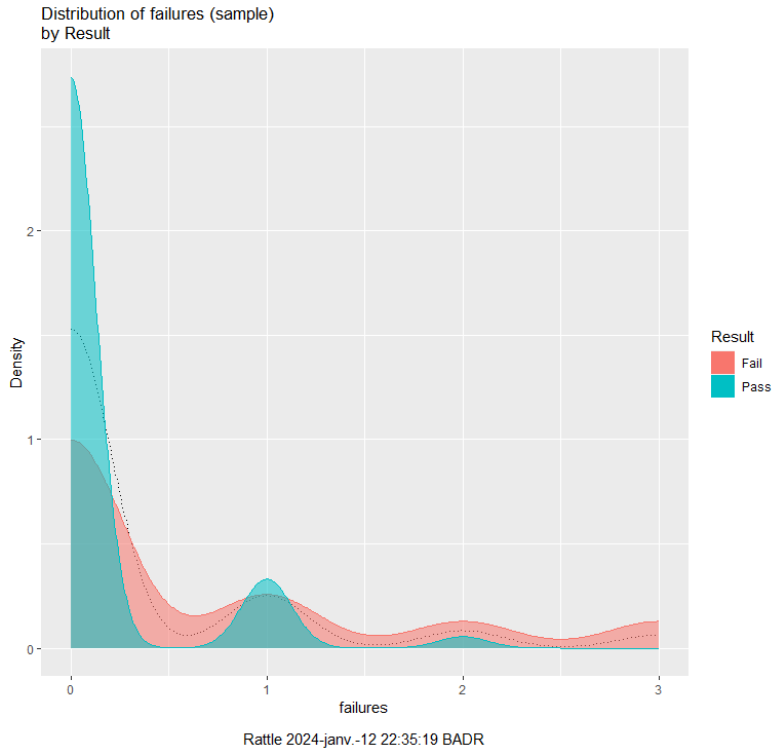


Interprétation

Les élèves qui ont réussi ont tendance à étudier plus que les élèves qui ont échoué.

Conclusion :

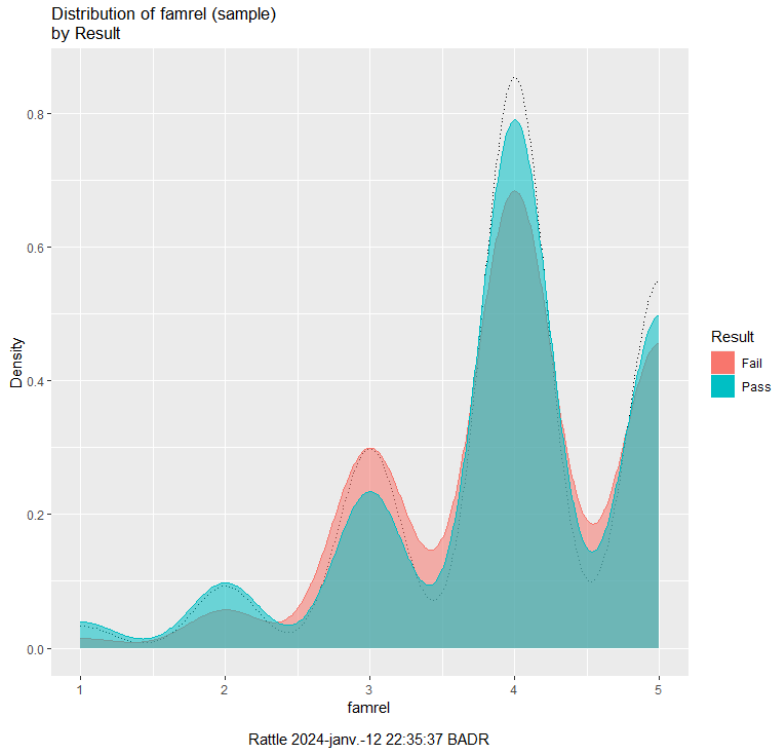
Le temps d'étude est un facteur important pour la réussite scolaire des enfants.



Interprétation

il existe une corrélation entre le nombre d'échecs de classe passés et les résultats aux examens. Cependant, il est important de noter que cette corrélation n'est pas causale. Il est possible que d'autres facteurs, tels que les capacités cognitives ou les motivations des élèves, puissent également influencer les résultats.

Conclusion : le nombre d'échecs de classe passés est un facteur important pour réussir aux examens, mais qu'il n'est pas le seul facteur.

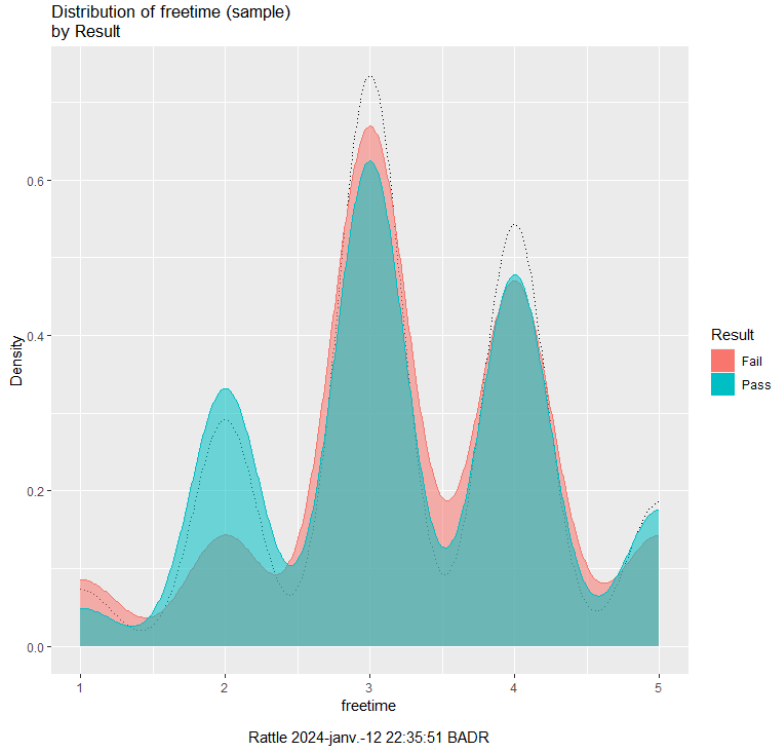


Interprétation

Les élèves qui ont des relations familiales plus positives ont tendance à avoir de meilleurs résultats aux examens.

Conclusion :

La qualité des relations familiales est un facteur important pour la réussite scolaire des enfants.

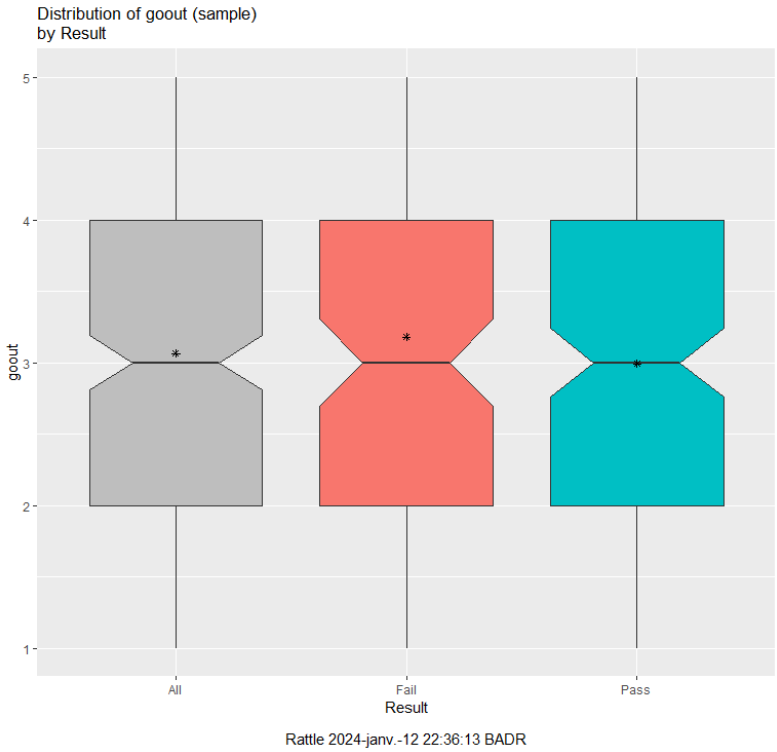


Interprétation

Les élèves qui ont réussi ont tendance à avoir plus de temps libre que les élèves qui ont échoué.

Conclusion:

La quantité de temps libre est un facteur qui peut contribuer à la réussite scolaire des élèves, mais elle n'est pas le seul facteur. D'autres facteurs, tels que les capacités cognitives, les motivations et les conditions de vie, peuvent également jouer un rôle.

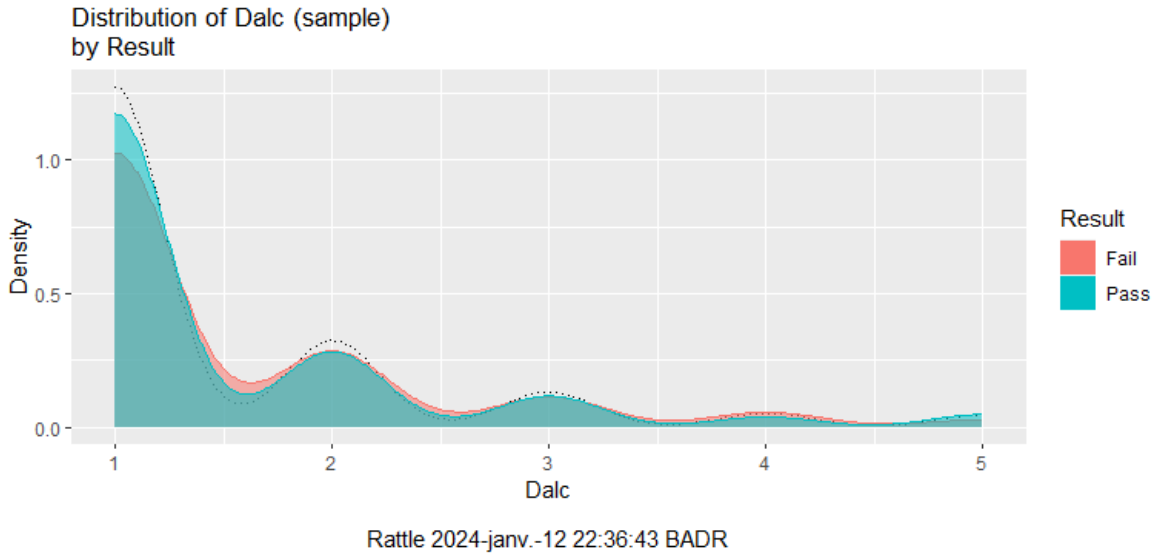


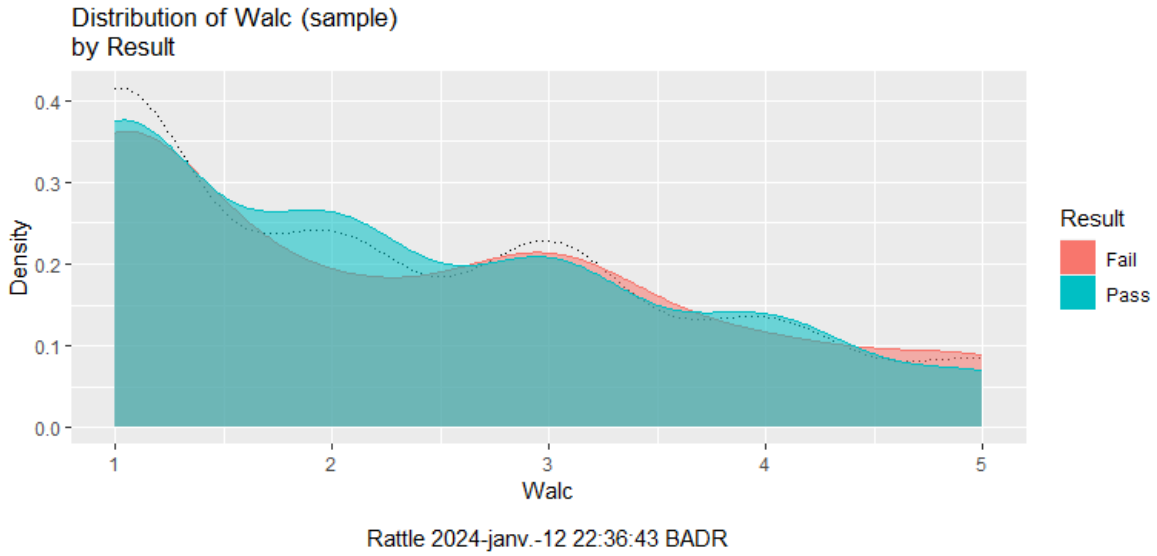
Interprétation

Les élèves qui ont réussi ont tendance à sortir avec des amis plus souvent que les élèves qui ont échoué.

Conclusion :

La fréquence des sorties avec des amis est un facteur qui peut contribuer à la réussite scolaire des élèves, mais elle n'est pas le seul facteur. D'autres facteurs, tels que les capacités cognitives, les motivations et les conditions de vie, peuvent également jouer un rôle.



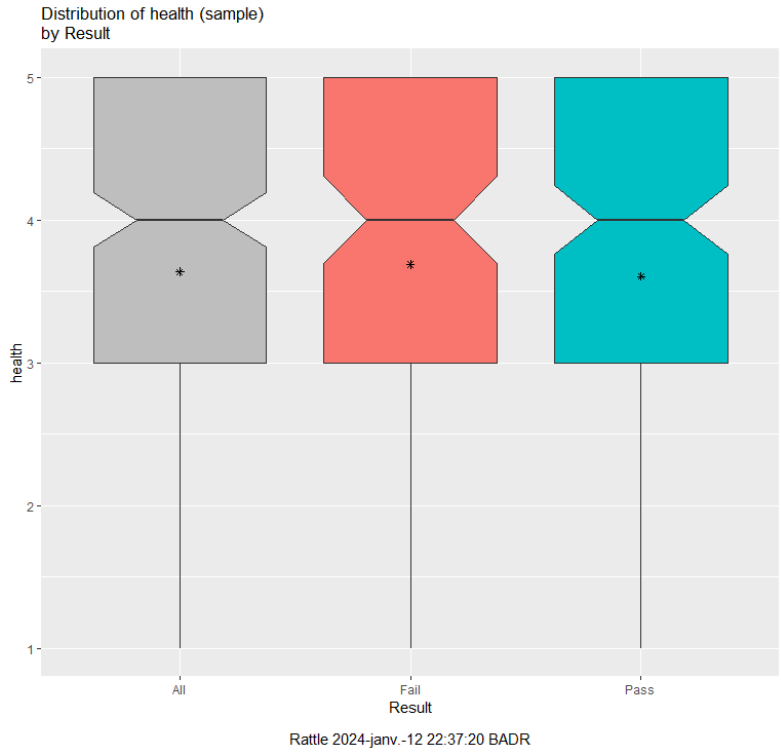


Interprétation

Les élèves qui ont réussi ont tendance à consommer moins d'alcool au travail et le week-end que les élèves qui ont échoué.

Conclusion :

La consommation d'alcool au travail et le week-end est un facteur qui peut contribuer à la réussite scolaire des élèves, mais elle n'est pas le seul facteur. D'autres facteurs, tels que les capacités cognitives, les motivations et les conditions de vie, peuvent également jouer un rôle.

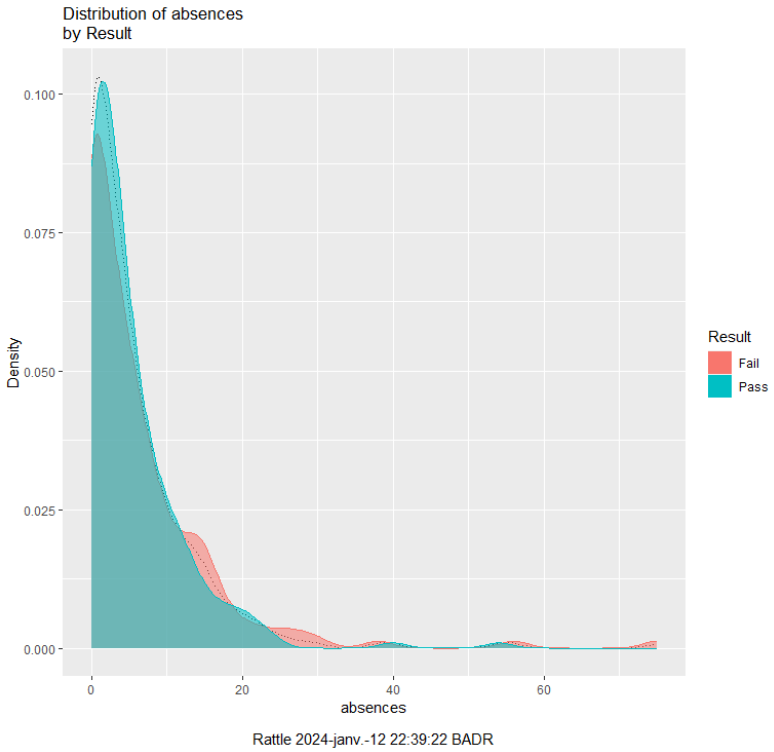


Interprétation

Les élèves qui ont réussi ont tendance à avoir un meilleur état de santé actuel que les élèves qui ont échoué.

Conclusion :

L'état de santé actuel est un facteur qui peut contribuer à la réussite scolaire des élèves, mais il n'est pas le seul facteur. D'autres facteurs, tels que les capacités cognitives, les motivations et les conditions de vie, peuvent également jouer un rôle.

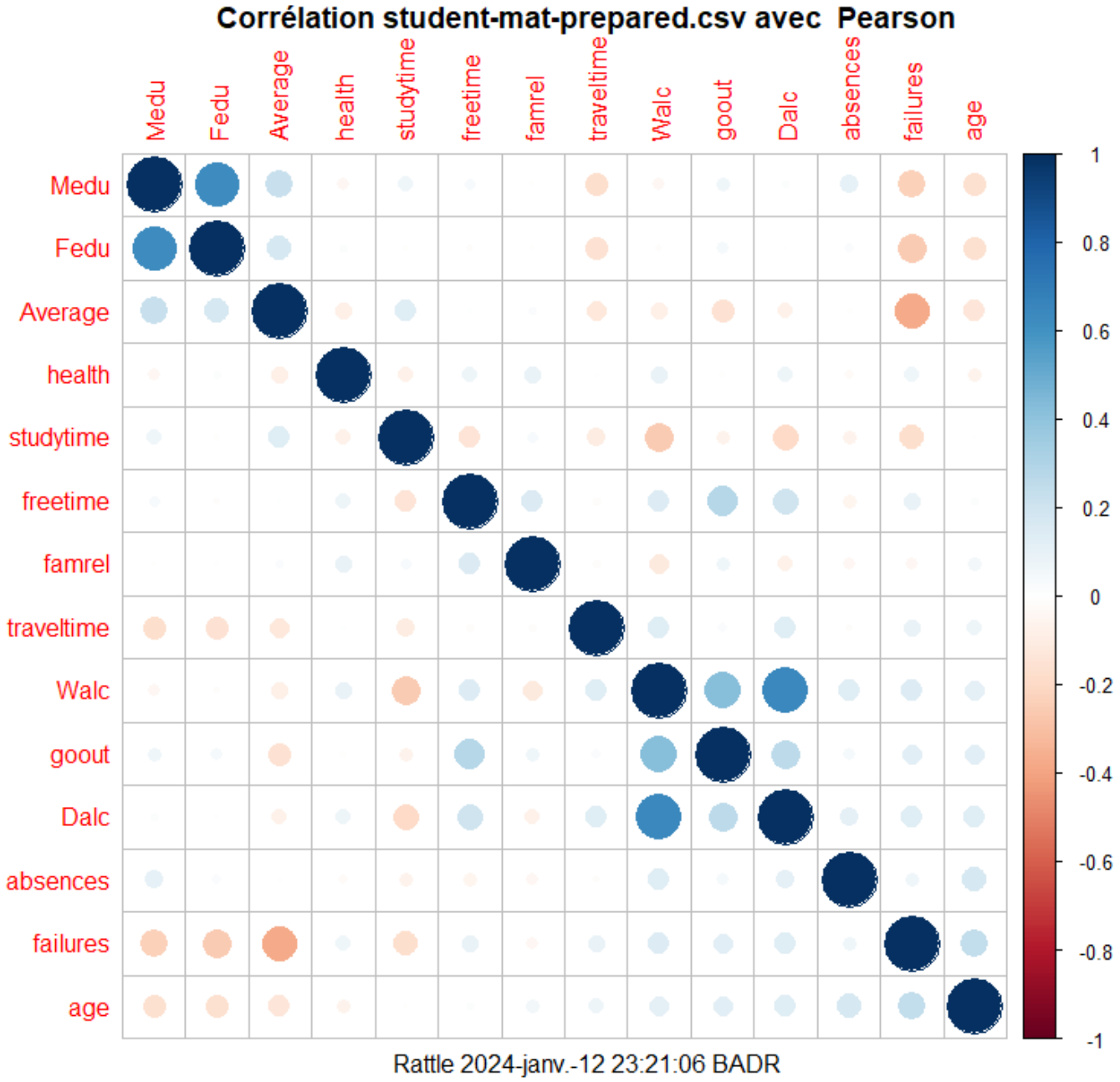


Interprétation

Les élèves qui ont réussi ont tendance à avoir moins d'absences scolaires que les élèves qui ont échoué.

Conclusion :

Le nombre d'absences scolaires est un facteur qui peut contribuer à la réussite scolaire des élèves, mais il n'est pas le seul facteur. D'autres facteurs, tels que les capacités cognitives, les motivations et les conditions de vie, peuvent également jouer un rôle.



Interprétation

La matrice de corrélation ci-dessous montre la corrélation entre les variables suivantes :

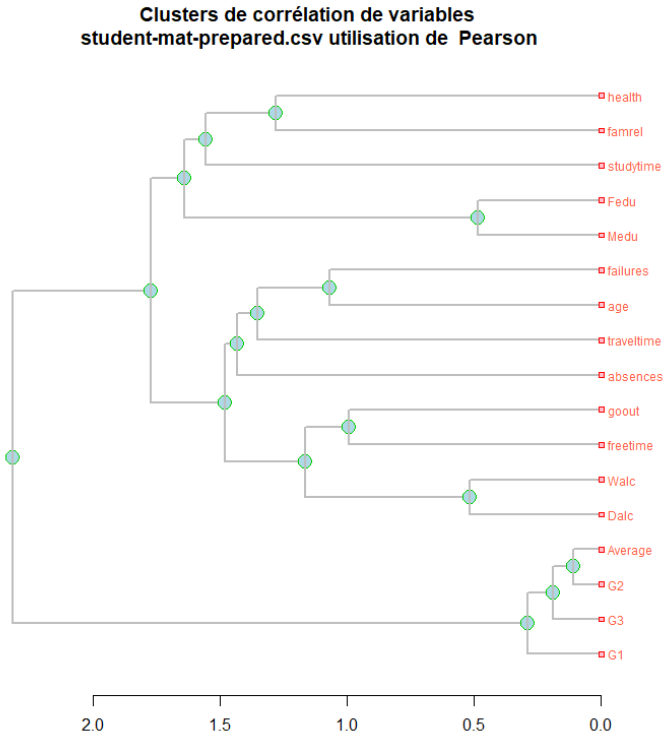
-Medu : Niveau d'éducation de la mère
-Fedu : Niveau d'éducation du père
-Average : Notes moyennes
-health : État de santé actuel
-studytime : Temps passé à étudier chaque jour
-failures : Nombre d'échecs scolaires
-absences : Nombre d'absences scolaires
-Walc : Consommation d'alcool le week-end
-goout : Fréquence des sorties avec des amis

Les valeurs de corrélation sont comprises entre -1 et 1. Une valeur de corrélation de 1 indique une corrélation positive parfaite, ce qui signifie que les deux variables augmentent ou diminuent ensemble. Une valeur de corrélation de -1 indique une corrélation négative parfaite, ce qui signifie que les deux variables augmentent ou diminuent l'une par rapport à l'autre. Une valeur de corrélation de 0 indique qu'il n'y a aucune corrélation entre les deux variables.

Il est important de noter que ces corrélations ne sont pas causales. Cela signifie que le fait qu'une variable soit corrélée à une autre ne signifie pas que la première variable cause la deuxième. Par exemple, le fait que l'âge et les résultats scolaires soient corrélés négativement ne signifie pas que l'âge est la cause des mauvais résultats scolaires. Il est possible que d'autres facteurs, tels que la motivation ou les compétences cognitives, jouent également un rôle.

Malgré cela, ces corrélations peuvent être utiles pour comprendre les facteurs qui influencent les résultats scolaires. Elles peuvent être utilisées pour développer des programmes d'intervention visant à améliorer les résultats scolaires des élèves.
En particulier, les résultats de cette étude suggèrent que les écoles devraient :
-Offrir des programmes de soutien aux élèves plus âgés :
-Promouvoir l'éducation des parents :
-Réduire le temps de trajet vers l'école :
-Encourager les élèves à travailler plus dur à l'école :
-Aider les élèves à gérer leur stress et leurs émotions :
-Créer un environnement familial positif :
-Offrir des activités extrascolaires saines :
-Prévenir la consommation d'alcool et de drogue :

Dans notre cas on va utiliser les variables corrélées pour prédire la note des

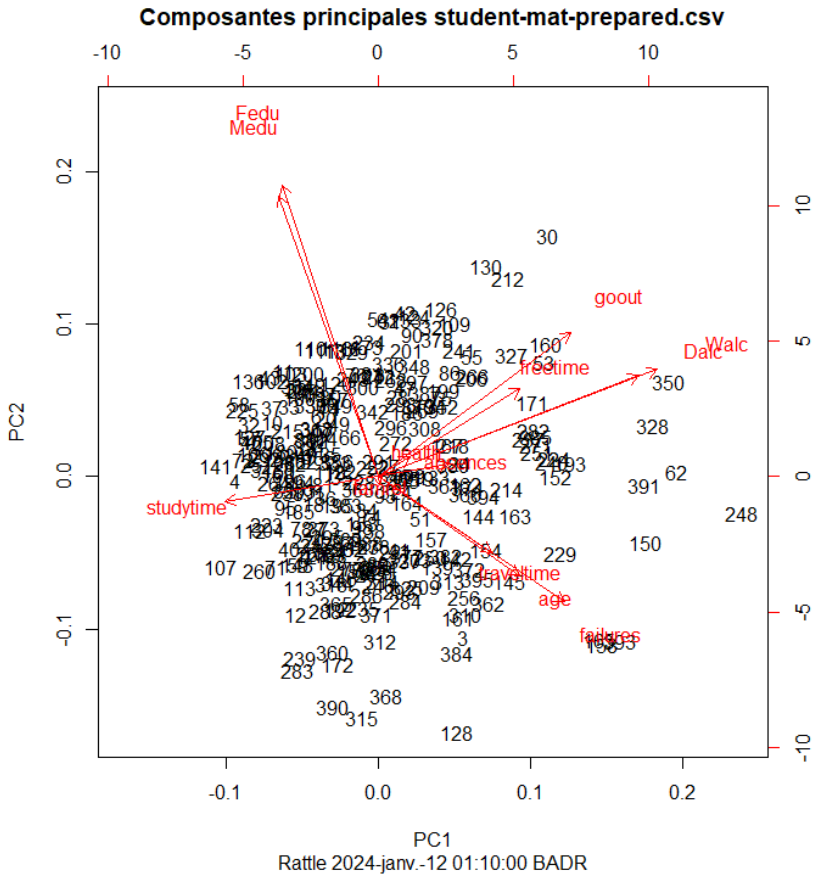


Interprétation

Corrélations globales :
La carte thermique montre les corrélations entre divers facteurs de performance des élèves.
Les corrélations les plus fortes semblent être entre :
Temps d'étude et échecs (négatif) : Les élèves qui étudient davantage ont tendance à avoir moins d'échecs.
Medu et Fedu (positif) : Les parents ayant un niveau d'éducation plus élevé ont tendance à avoir des enfants ayant un niveau d'éducation plus élevé.
Walc et goout (positif) : Les élèves qui consomment plus d'alcool le week-end ont tendance à sortir plus souvent avec des amis.
Corrélations spécifiques :
Santé et échecs (négatif) : Les élèves en meilleure santé ont tendance à avoir moins d'échecs.
Absences et échecs (positifs) : Les élèves qui ont plus d'absences ont tendance à avoir plus d'échecs.

Moyenne et temps d'étude (positif) : Les élèves qui étudient davantage ont tendance à avoir des notes moyennes plus élevées.
Moyenne et sortie (négative) : Les élèves qui sortent plus souvent ont tendance à avoir des notes moyennes plus faibles.
Il est important de noter que :

Les corrélations n'impliquent pas la causalité. Ce n'est pas parce que deux variables sont corrélées que l'une cause l'autre.
Ces corrélations sont basées sur un ensemble de données spécifique et peuvent ne pas être généralisables à d'autres populations.



Interprétation

Dans ce cas, les deux premières composantes principales expliquent environ 70 % de la variance totale des données. La première composante principale (PC1) est positivement corrélée avec les variables suivantes :

studytime (temps passé à étudier)
failures (nombre d'échecs)
absences (nombre d'absences)
La deuxième composante principale (PC2) est positivement corrélée avec les variables suivantes :

Medu (niveau d'éducation de la mère)
Fedu (niveau d'éducation du père)
Walc (consommation d'alcool le week-end)
En termes simples, la PC1 semble refléter la réussite scolaire globale des élèves, tandis que la PC2 semble refléter les facteurs socio-économiques des élèves.

Voici une interprétation plus détaillée des résultats :

PC1

La PC1 est positivement corrélée avec les variables suivantes :

studytime (temps passé à étudier)

Cela suggère que les élèves qui passent plus de temps à étudier ont tendance à avoir de meilleurs résultats scolaires. Cela est probablement dû au fait que les élèves qui étudient plus ont plus de chances de comprendre les concepts et les informations enseignés en classe.

failures (nombre d'échecs)
Cela suggère que les élèves qui ont plus d'échecs ont tendance à avoir de moins bons résultats scolaires. Cela est probablement dû au fait que les élèves qui ont plus d'échecs ont plus de chances de ne pas comprendre les concepts et les informations enseignés en classe.

absences (nombre d'absences)
Cela suggère que les élèves qui ont plus d'absences ont tendance à avoir de moins bons résultats scolaires. Cela est probablement dû au fait que les élèves qui ont plus d'absences ont plus de chances de manquer d'informations importantes enseignées en classe.

En conclusion, la PC1 semble refléter la réussite scolaire globale des élèves. Les élèves qui passent plus de temps à étudier, qui ont moins d'échecs et qui ont moins d'absences ont tendance à avoir de meilleurs résultats scolaires.

PC2

La PC2 est positivement corrélée avec les variables suivantes :

Medu (niveau d'éducation de la mère)

Fedu (niveau d'éducation du père)

Walc (consommation d'alcool le week-end)

Medu (niveau d'éducation de la mère)

Fedu (niveau d'éducation du père)

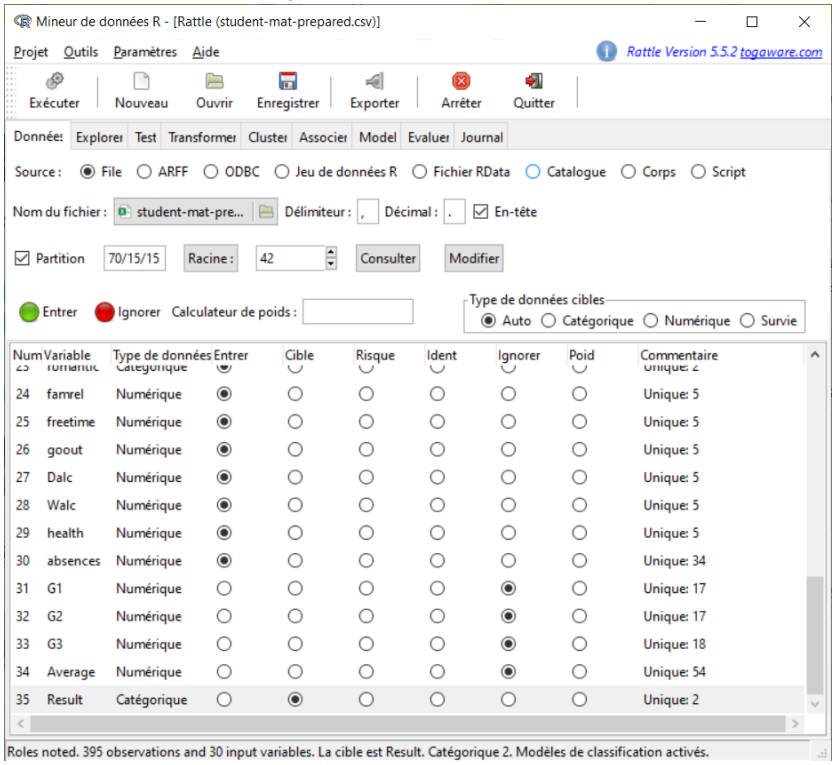
Cela suggère que les élèves dont les parents ont un niveau d'éducation élevé ont tendance à avoir de meilleurs résultats scolaires. Cela est probablement dû au fait que les enfants dont les parents ont un niveau d'éducation élevé ont plus de chances de bénéficier d'un environnement familial stimulant, d'un accès à des ressources éducatives et d'un soutien financier pour l'éducation.

Walc (consommation d'alcool le week-end)
Cela suggère que les élèves qui consomment plus d'alcool le week-end ont tendance à avoir de moins bons résultats scolaires. Cela est probablement dû au fait que la consommation d'alcool peut nuire aux capacités cognitives et à la motivation des élèves.

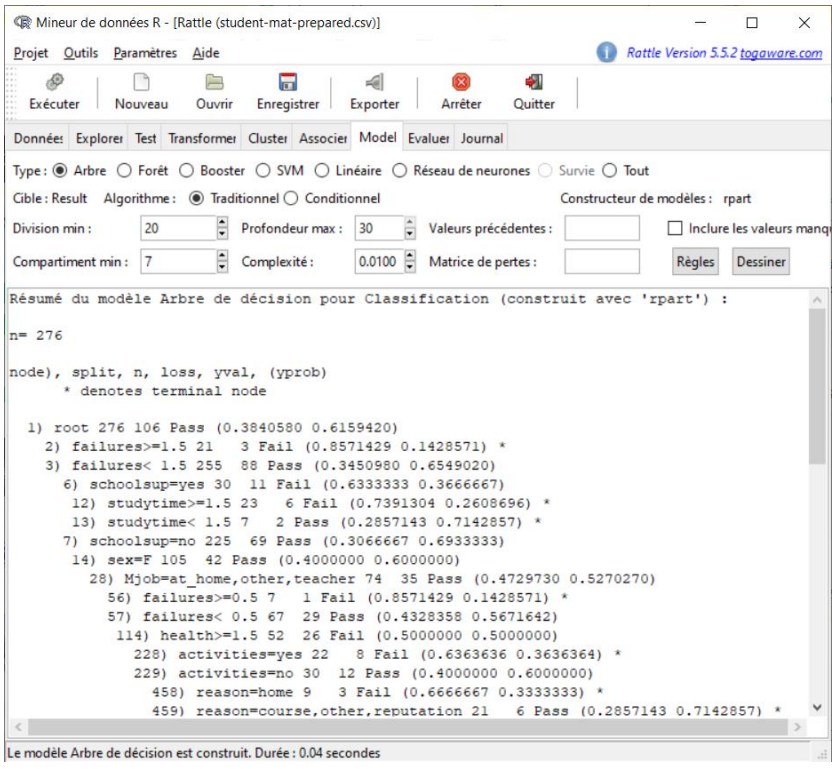
En conclusion, la PC2 semble refléter les facteurs socio-économiques des élèves. Les élèves dont les parents ont un niveau d'éducation élevé et qui consomment moins d'alcool ont tendance à avoir de meilleurs résultats scolaires.

Créer un model : arbre decision

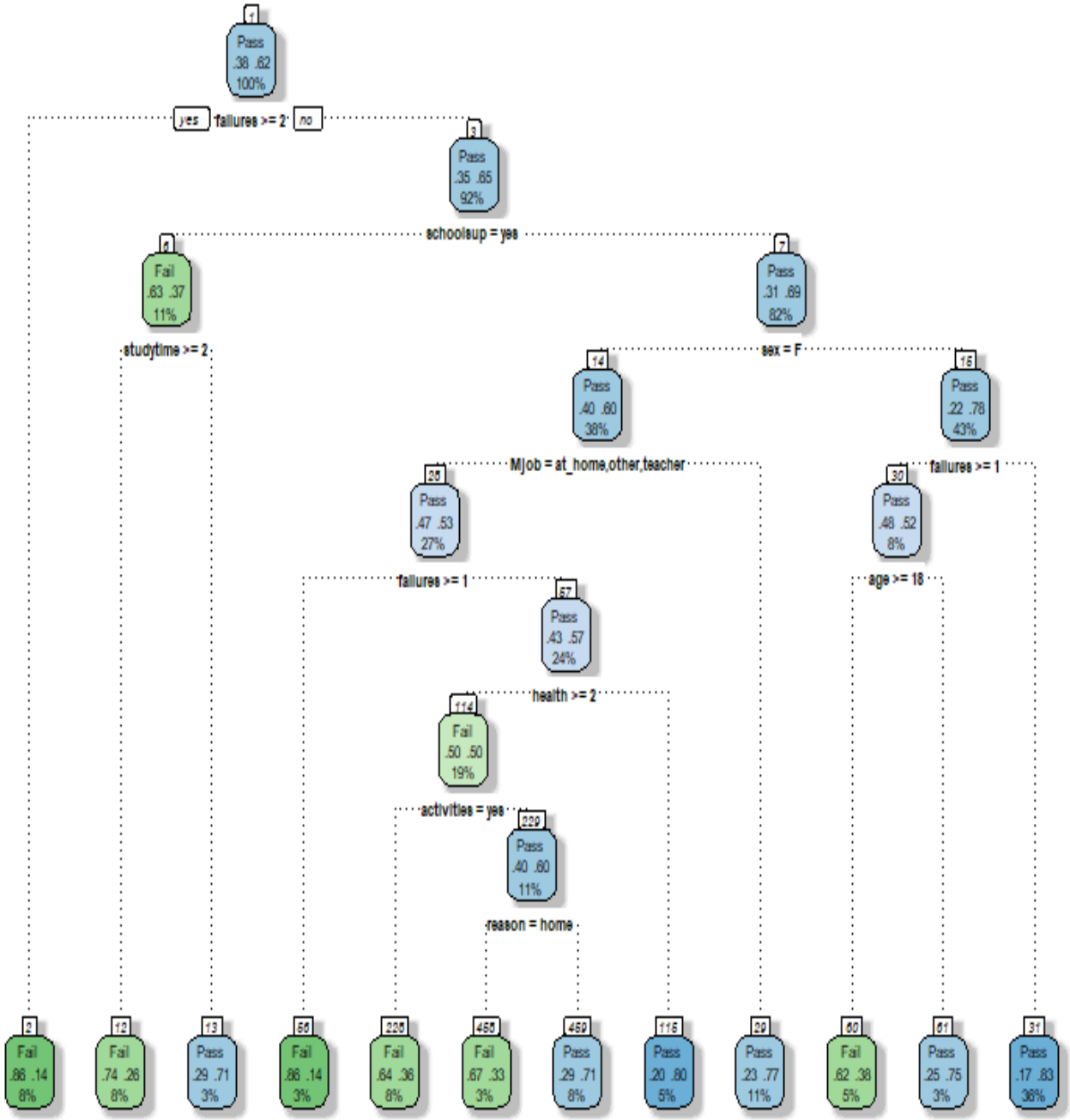
Charger la base comme suit et cliquer sur executer



Aller à model choisisez Arbe et definissez les paramèteres voulu quliquer sur exécuter



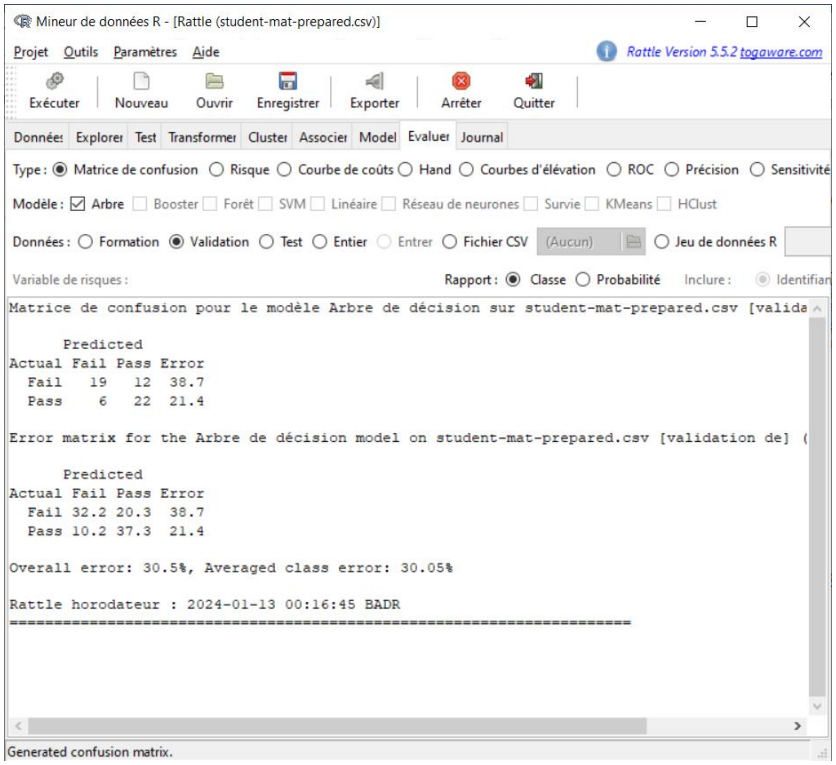
Arbre de décision student-mat-prepared.csv $ Result



Rattle 2024-janv.-12 01:10:34 BADR

Evaluation du model

Allez a evaluer



Matrice de confusion pour le modèle Arbre de décision sur student-mat-prepared.csv [validation de] (comptes) :

```
      Predicted
Actual Fail Pass Error
  Fail   19   12  38.7
  Pass    6   22  21.4
```

Error matrix for the Arbre de décision model on student-mat-prepared.csv [validation de] (proportions):

```
      Predicted
Actual Fail Pass Error
  Fail 32.2 20.3  38.7
  Pass 10.2 37.3  21.4
```

Overall error: 30.5%, Averaged class error: 30.05%

Rattle horodateur : 2024-01-13 00:16:45 BADR

Interprétation

Performances du modèle :
Erreur globale : le modèle prédit correctement 70 % des cas (erreur de 100 % à 30,5 %).
Erreur de classe moyenne : les taux d'erreur du modèle pour prédire « Échec » et « Réussite » sont relativement similaires, avec une moyenne de 30,05 %.
Considérations :

- -Qualité des données : La précision du modèle dépend de la qualité des données sur lesquelles il a été formé.
- -Déséquilibre de classe : S'il y a beaucoup plus d'élèves dans une classe (échec ou réussite) que dans l'autre, cela peut affecter les taux d'erreur.
- -Réglage du modèle : L'ajustement des paramètres de l'arbre de décision peut améliorer ses performances.
- -Modèles alternatifs : L'exploration de différents algorithmes d'apprentissage automatique pourrait donner de meilleurs résultats.