

Projet final 1

Titre : *Analyse de performance étudiante*

Nom complet : Frantzdy DORILUS

Date : 11 Décembre 2025

Résumé

Ce rapport présente une analyse complète des performances académiques d'un groupe d'étudiants sur deux semestres consécutifs. L'étude a pour objectif principal d'identifier les facteurs associés à la réussite scolaire et de proposer des recommandations concrètes pour améliorer les résultats.

Les données utilisées incluent la provenance(ville), les notes par matière, la moyenne générale, la filière, le genre, l'âge, le nombre d'heures d'étude hebdomadaires et le taux d'absences. L'approche méthodologique combine des analyses descriptives, des tests statistiques (ANOVA, tests T indépendants et appariés) et des visualisations graphiques (boxplots, scatter plots, diagrammes de corrélation) pour explorer les relations entre les variables et la performance des étudiants.

Les résultats montrent que les performances varient significativement selon la filière et non selon le genre et que l'évolution des moyennes entre le semestre 1 et le semestre 2 reste globalement stable mais affiche à la fois des progressions et des stagnations selon les groupes. Les heures d'étude et le nombre d'absences apparaissent comme des facteurs déterminants : un engagement régulier dans les études est positivement corrélé à la performance tandis que des absences fréquentes sont associées à une baisse des résultats.

Enfin, le rapport propose des recommandations concrètes visant à optimiser le temps d'étude, réduire les absences et accompagner les étudiants à risque contribuant ainsi à améliorer la réussite académique globale.

Introduction

La réussite académique des étudiants constitue un enjeu majeur pour les établissements d'enseignement supérieur, tant pour l'amélioration de la qualité de la formation que pour le développement du capital humain. Comprendre les facteurs qui influencent les performances scolaires permet non seulement d'optimiser les méthodes pédagogiques, mais aussi de mettre en place des dispositifs d'accompagnement adaptés aux besoins des étudiants.

Dans le contexte de cette étude, nous nous intéressons à un groupe d'étudiants dont les profils démographiques, comportementaux et académiques ont été suivis sur deux semestres consécutifs. Les variables analysées incluent, entre autres, la filière, le genre, l'âge, les notes obtenues dans différentes matières, le nombre d'heures d'étude hebdomadaires et le nombre d'absences. Ces éléments permettent de dresser un portrait global des étudiants et d'évaluer l'impact de différents facteurs sur leur performance.

Objectifs de l'étude :

Les objectifs principaux de cette analyse sont :

1. **Caractériser les profils démographiques et comportementaux des étudiants** : identifier la répartition par filière, genre, tranche d'âge, et observer les variations de performance selon ces caractéristiques.
2. **Analyser l'évolution des performances entre les deux semestres** : détecter les tendances générales et les changements significatifs dans les résultats académiques

3. **Identifier les facteurs associés à la réussite académique** : mesurer l'effet des heures d'étude et des absences, et déterminer les seuils optimaux pour maximiser les performances.
4. **Formuler des recommandations concrètes** : proposer des actions ciblées pour améliorer les résultats scolaires et réduire le risque d'échec.

Cette étude vise donc à fournir une compréhension fine des déterminants de la réussite scolaire et à orienter les stratégies pédagogiques et administratives pour soutenir les étudiants à risque tout en valorisant les pratiques efficaces déjà en place.

Méthodologie

Cette section décrit la structure des données utilisées, les variables étudiées et les méthodes statistiques mobilisées pour répondre aux questions de recherche. L'analyse repose sur un ensemble d'informations collectées auprès d'étudiants inscrits dans différentes filières et couvre leurs caractéristiques sociodémographiques, leurs comportements d'étude et leurs performances académiques sur deux semestres.

Description et catégorisation des variables

Les tableaux suivant présentent l'ensemble des variables utilisées dans l'étude, leur type ainsi que leur signification.

A. Variables d'identification

Variable	Type	Description
id_etudiant	<i>Caractère</i>	Identifiant unique de l'étudiant (ex. : ETU001–ETU200)
nom	<i>Caractère</i>	Nom ou étiquette textuelle de référence (non utilisé dans les analyses)

B. Variables sociodémographiques

Variable	Type	Description
age	<i>Numérique</i>	Âge de l'étudiant en années
genre	<i>Facteur</i>	Sexe de l'étudiant (Homme / Femme)
ville	<i>Caractère</i>	Commune d'origine ou de résidence
filiere	<i>Facteur</i>	Filière d'études (Informatique, Economie, Mathématique, Physique etc.)

C. Variables comportementales

Variable	Type	Description
heures_etude_semaine	Numérique	Nombre d'heures d'étude hebdomadaires
nb_absences_s1	Numérique	Nombre d'absences au Semestre 1
nb_absences_s2	Numérique	Nombre d'absences au Semestre 2

D. Variables académiques

Variable	Type	Description
moyenne_s1	Numérique	Moyenne générale au Semestre 1
moyenne_s2	Numérique	Moyenne générale au Semestre 2

E. Notes Académiques

Semestre 1 : 5 matières

- note_math_s1 : Note de mathématiques (0-20)
- note_info_s1 : Note d'informatique (0-20)
- note_physique_s1 : Note de physique (0-20)
- note_economie_s1 : Note d'économie (0-20)
- note_anglais_s1 : Note d'anglais (0-20)

Semestre 2 : 5 matières

- note_math_s2 : Note de mathématiques (0-20)
- note_info_s2 : Note d'informatique (0-20)
- note_physique_s2 : Note de physique (0-20)
- note_economie_s2 : Note d'économie (0-20)
- note_anglais_s2 : Note d'anglais (0-20)

2. Approche analytique

Pour répondre aux objectifs et questions de recherche, nous avons appliqué plusieurs méthodes complémentaires :

A. Analyses descriptives

- Statistiques de base : moyenne, médiane, Ecart-type , skewness (asymétrie), minimum/maximum.
- Distribution des notes et moyennes par semestre.
- Répartition des étudiants selon la filière, le genre et la ville.

B. Analyses comparatives

- Test t apparié : comparaison des moyennes entre S1 et S2.
- Tests t indépendants : comparaison selon le genre.
- ANOVA : comparaison des moyennes entre filières.

C. Analyses de corrélation

- heures d'étude et moyennes
- absences et performances

D. Visualisations

- Histogramme (moyennes générale)
- Boxplot (heures vs performance, absences vs performance)
- Scatterplot (distribution des filières, villes)
- Line charts (évolution S1 → S2)
- Heatmap des corrélations

3. Outils logiciels

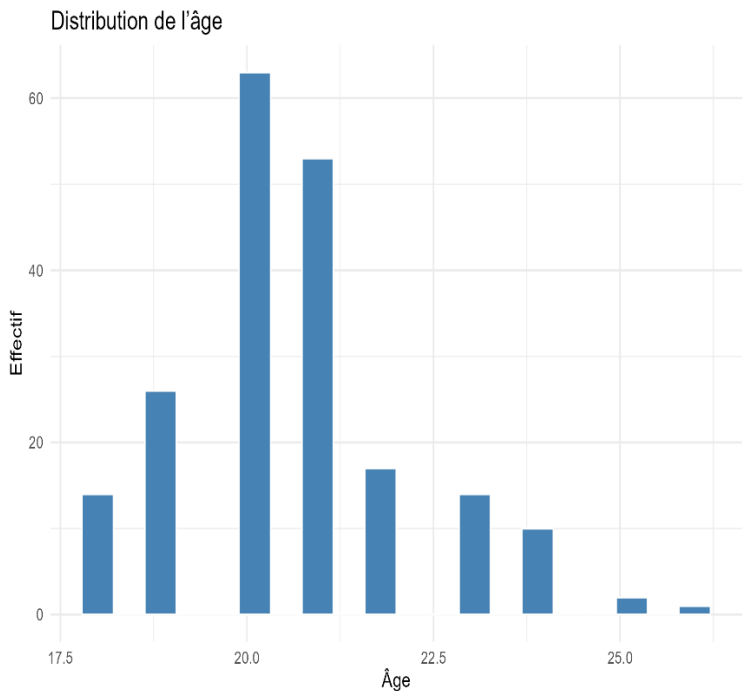
L'analyse a été entièrement réalisée avec **R** en utilisant principalement les packages :

- **tidyverse** : manipulation et visualisation
- **ggplot2** : graphiques
- **dplyr** et **tidyr** : nettoyage et transformation
- **moments** : calcul du coefficient d'asymétrie
- **readr** / **readxl** : importation des données

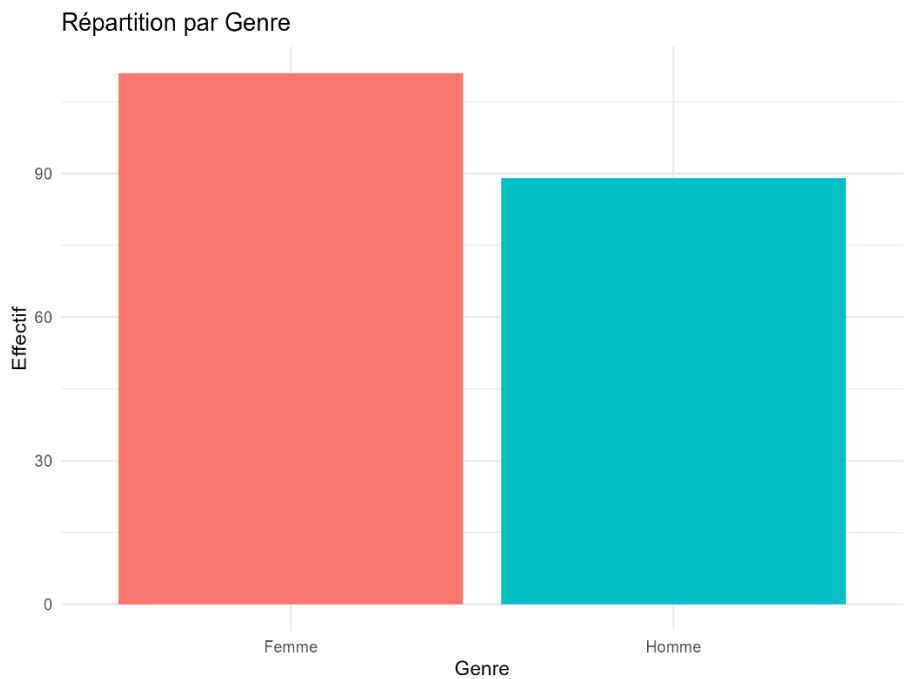
4) Résultats

4.1 Profil démographique des étudiants

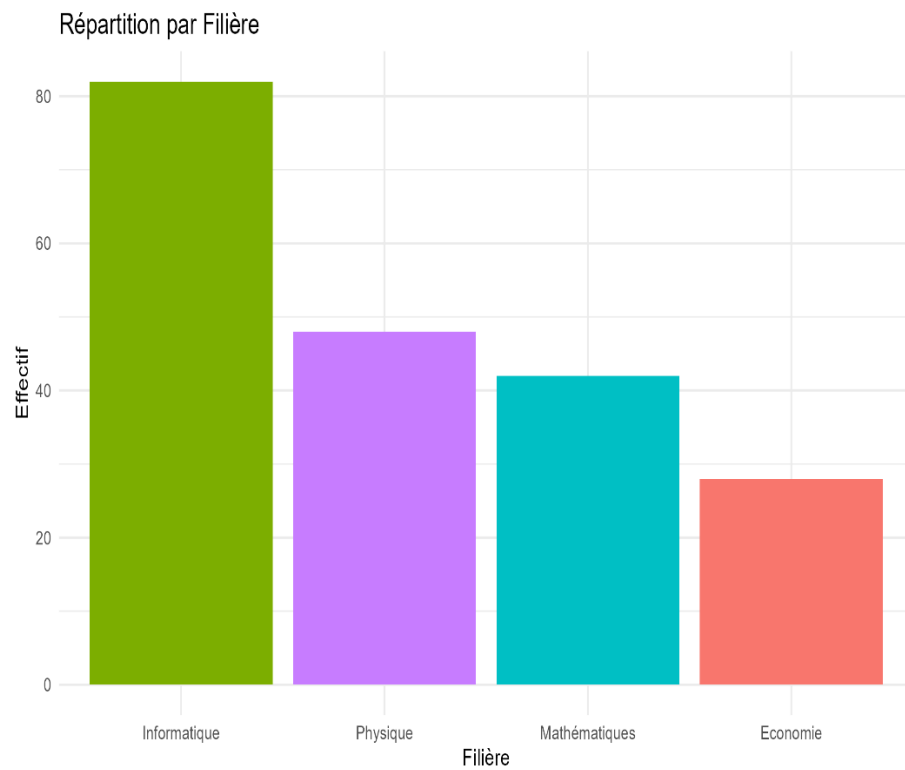
4.1.1 Distribution des étudiants selon l'âge



4.1.2 Répartition par genre

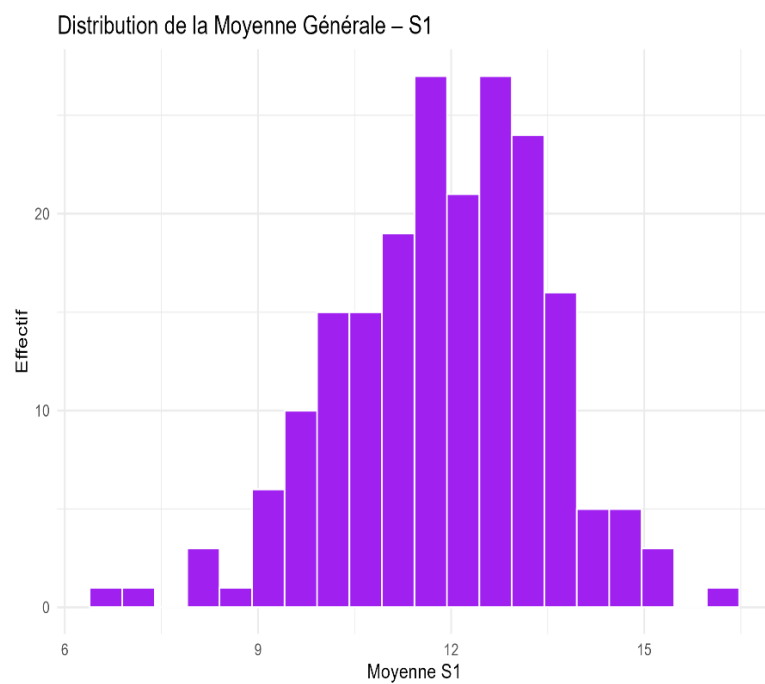


4.1.3 Répartition par filière

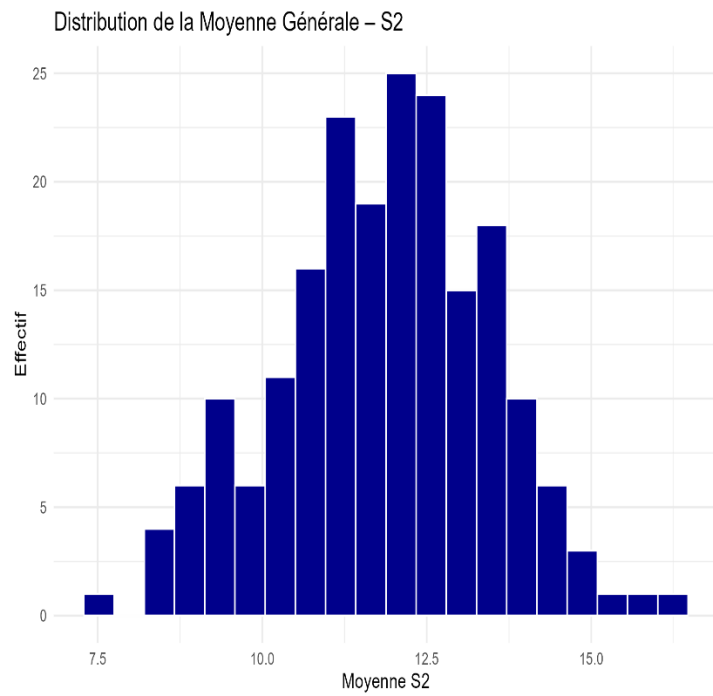


4.2 Analyse des performances académiques

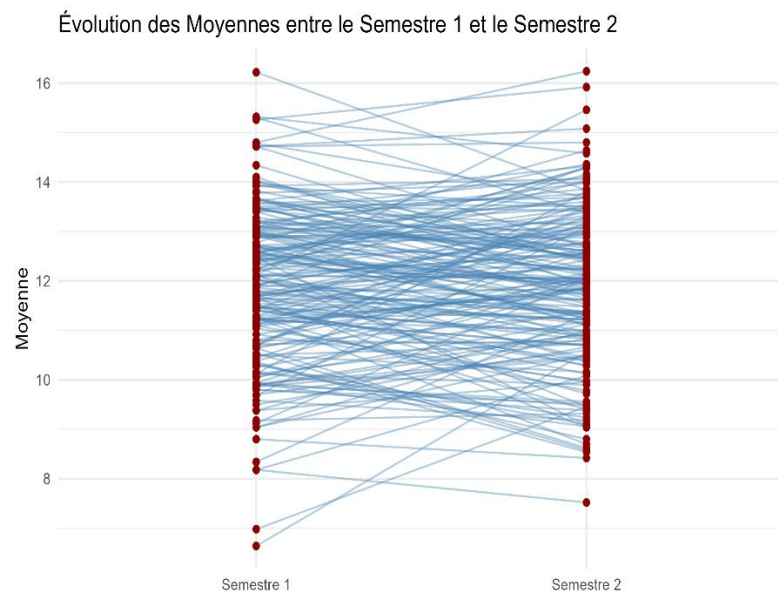
4.2.1 Moyenne générale S1



4.2.2 Moyenne générale S2

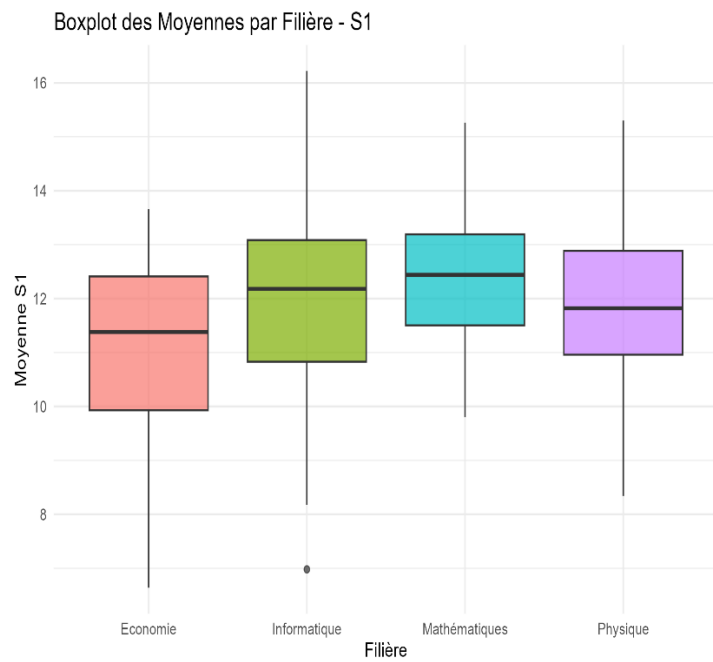


4.2.3 Evolution des moyennes

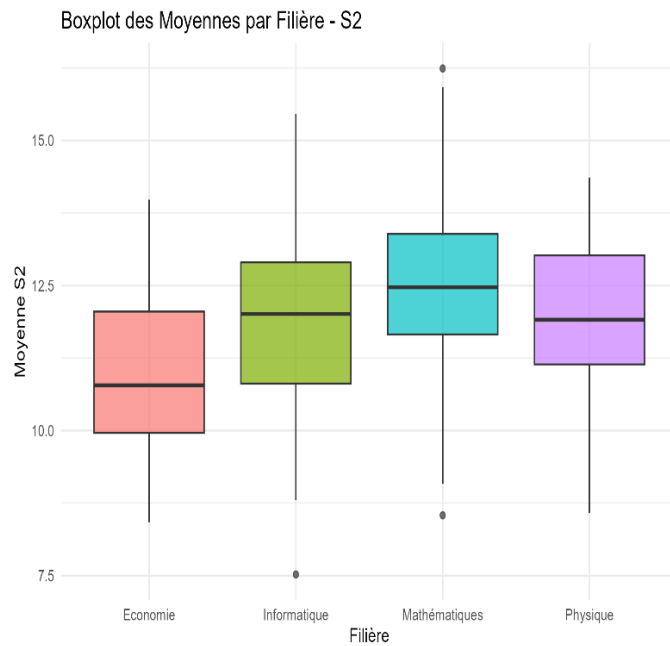


4.3 Performance par groupes

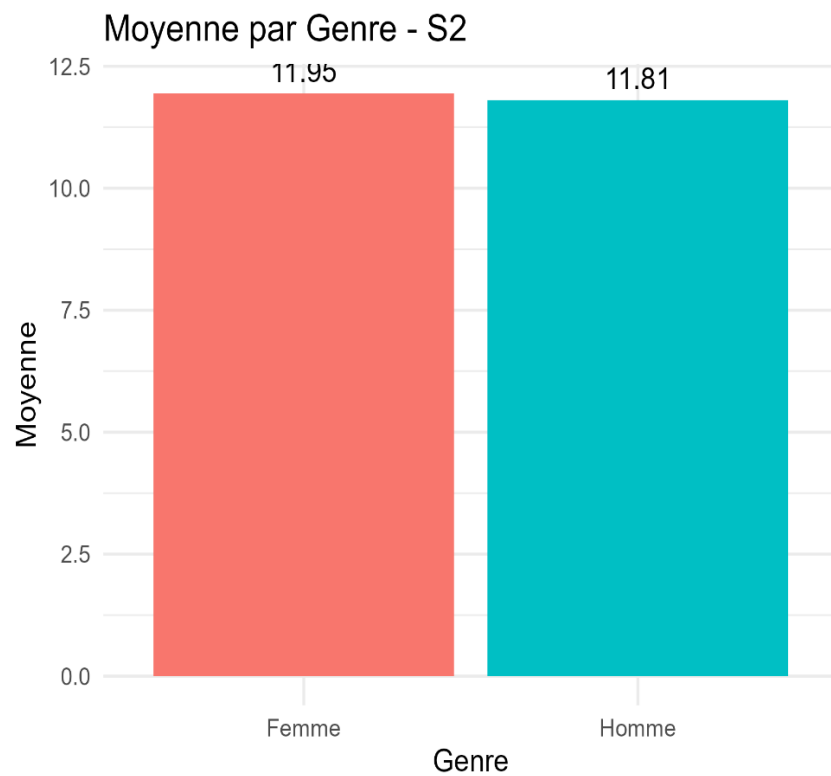
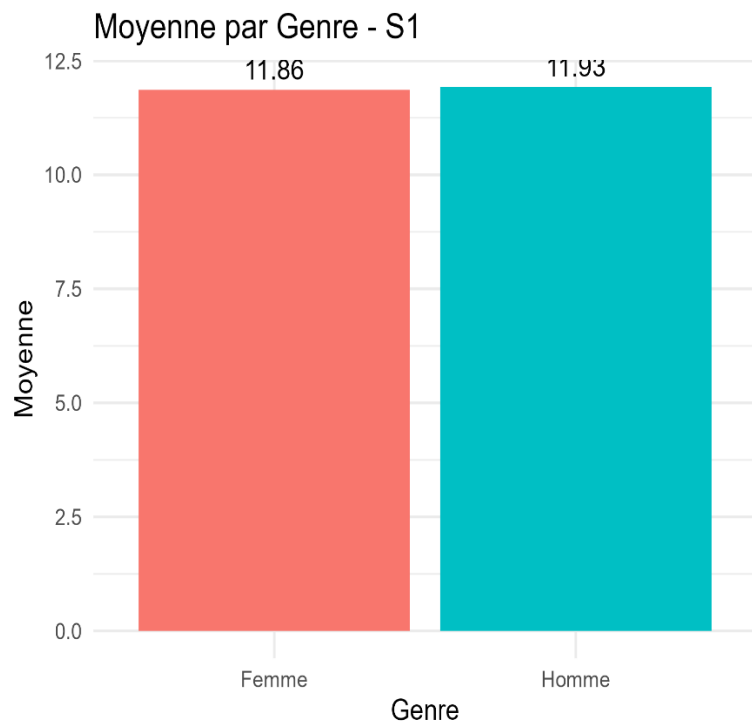
4.3.1 Performance par filière S1



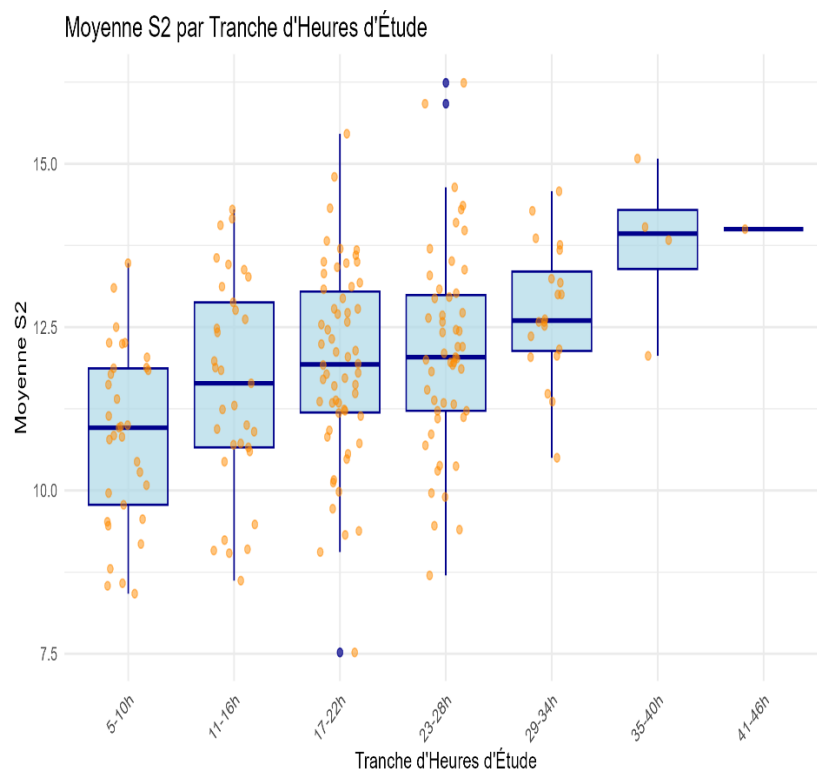
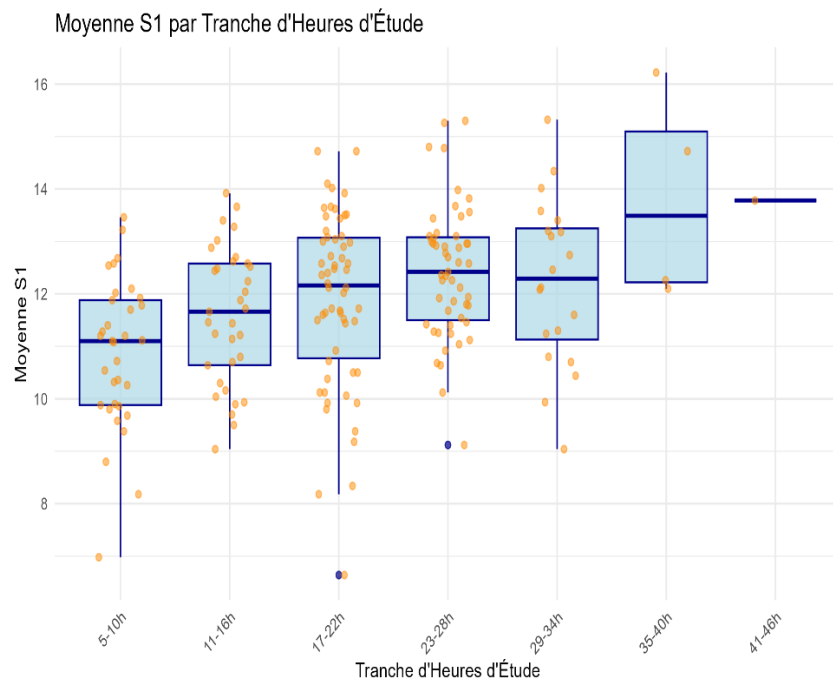
4.3.2 Performance par filière S2



4.3.3 Analyse de la performance par genre

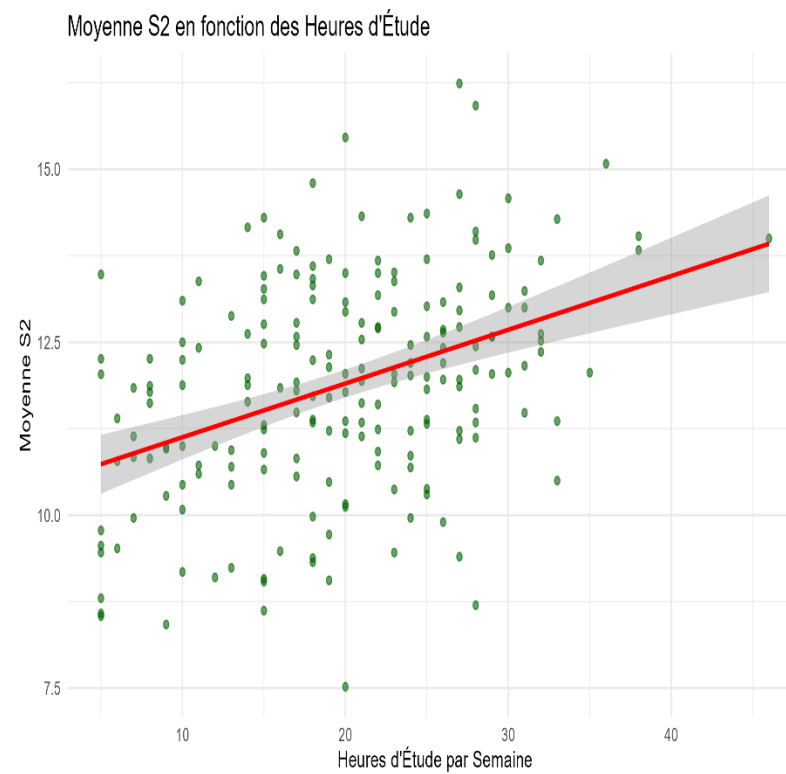
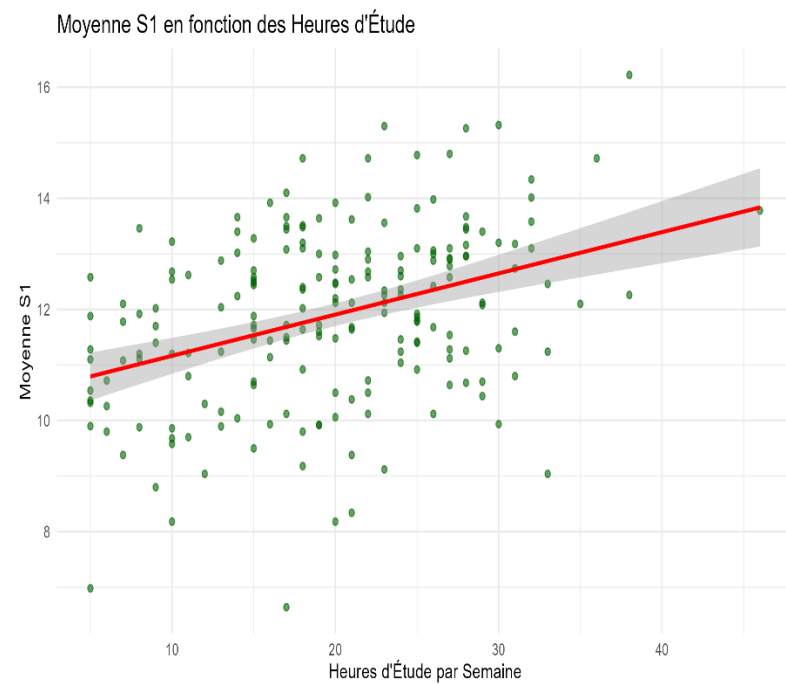


4.3.4 Analyse de la performance par tranche d'heure (S1 et S2)

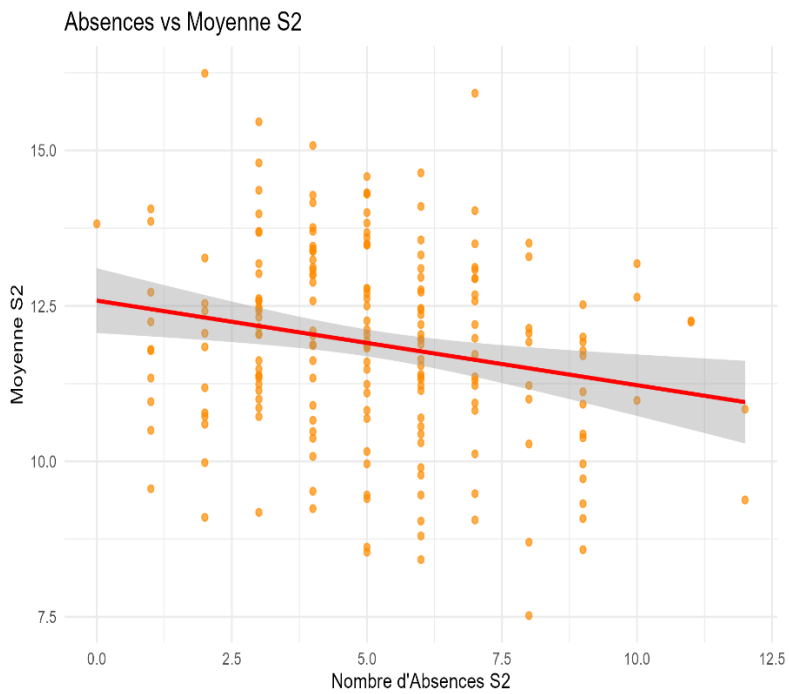
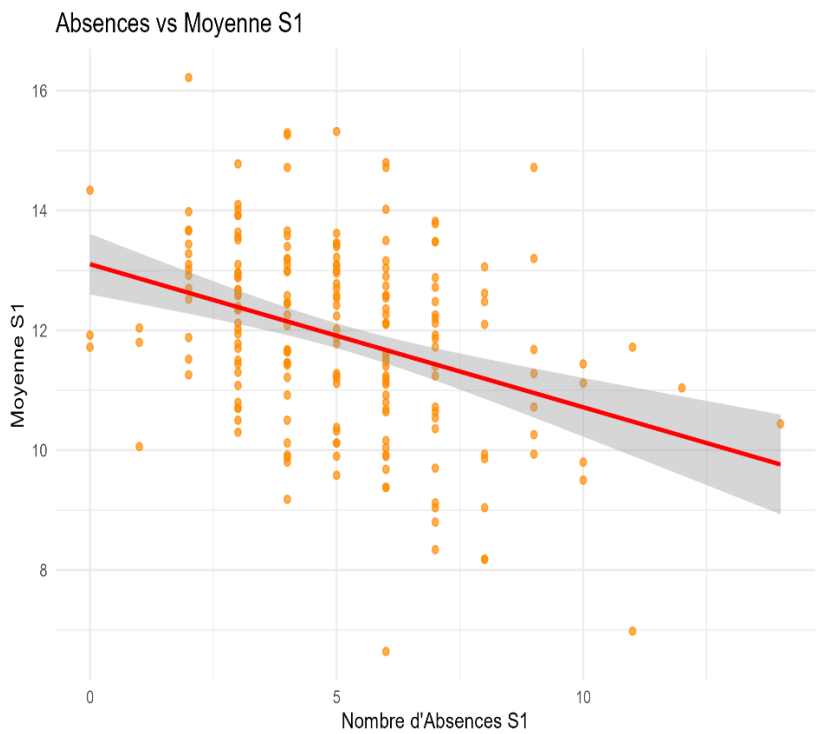


4.4 Analyse des facteurs de réussite

4.4.1 Impact des heures d'étude



4.4.2 Impact des absences



5) Discussions

1. Profil démographique des résultats.

L'échantillon comprend 200 étudiants avec un âge moyen de 20,7 ce qui veut dire si l'on fait la distribution de manière à ce que les étudiants aient le même âge ce serait 20,7 ans. Une médiane de 20 ans veut dire que 50% des étudiants ont un âge supérieur à 20 ans et 50% ont un âge inférieure à 20 ans. Un écart type relativement faible (1,57) qui représente une faible dispersion. On peut donc conclure que l'âge des étudiants est proche de l'âge moyen. L'âge minimum est de 18 ans et le maximum de 26 ans, ce qui indique une population homogène et typique d'un public universitaire jeune sans présence d'âges extrêmes ou atypiques. Dans cette population les femmes représentent 55,5%.

La répartition académique révèle une structure clairement déséquilibrée entre les filières. La filière Informatique constitue presque la moitié de l'échantillon ce qui peut influencer les tendances globales observées dans les performances.

2. Performance académique

a) Globale

Les histogrammes représentent la distribution des moyennes générales pour les semestres S1 et S2. Ils mettent en évidence une répartition relativement similaire entre les deux périodes avec un taux global de réussite respectif 87% et 86,5%. Dans l'ensemble, les moyennes se concentrent autour de 12/20 ce qui confirme les valeurs observées dans les statistiques descriptives (moyenne et médiane de 11.9 et 12).

Cette distribution présente une forme symétrique et légèrement resserrée ce qui indique une faible dispersion des résultats autour de la moyenne générale (écart-type d'environ 1.58 pour les deux semestres). Autrement dit, la majorité des étudiants ont obtenu des performances proches les uns des autres.

b) Par genre

Cette section examine les différences de performance entre les étudiants femmes et hommes pour les deux semestres, à l'aide d'un test t de Student indépendant.

Comparaison des Moyenne – Semestre (1 et 2)

- **Variable dépendante** : Moyenne générale (S1 et S2)
- **Variable indépendante** : Genre (Femme / Homme)

Démarche du test t

1. Hypothèses

- H_0 : La moyenne générale (S1 et S2) des femmes est égale à celle des hommes
$$\mu_{Femme} = \mu_{Homme}$$
- H_1 : La moyenne générale (S1 et S2) des femmes est différente de celle des hommes

$$\mu_{Femme} \neq \mu_{Homme}$$

2. Seuil de signification: $\alpha = 0.05$

3. Résultat

- p-value S1 = 0.762
- p-value S2 = 0.537

4. Règle de décision:

- Si p-value < 0.05 on rejette H0
- Sinon on ne rejette pas H0

Conclusion : p-value (S1 et S2) > 0.05, on ne rejette pas H0. Il n'y a donc aucune différence significative de performance entre femmes et hommes dans les deux semestres.

c) Par filière

Les résultats montrent que la filière Mathématiques apparaît comme la plus performante dans les deux semestres tandis que la filière Économie affiche les résultats les plus faibles. Pour déterminer si ces différences observées sont statistiquement significatives, un test ANOVA (Analysis Of Variance) à un facteur a été réalisé.

• Hypothèses du test

- **H0** (hypothèse nulle) : les moyennes S1 sont identiques entre toutes les filières aucune différence de performance selon la filière
- **H1** (hypothèse alternative) : au moins une filière présente une moyenne significativement différente

• Seuil de signification

- $\alpha = 5 \%$

• Résultat du test

- p-value (semestre 1) = 0.0129
- p-value (semestre 2) = 0.0038

• Règle de décision

- Si p-value < α , on rejette H0.
- Ici : p-value (S1 et S2) < 0.05 on rejette H0

Conclusion : Il existe une différence significative de performance académique selon la filière dans les deux semestres. La filière Mathématiques se démarque comme la plus performante tandis que la filière Économie présente des résultats significativement plus faibles

d) Par tranche d'âge

Cette section examine si la performance académique diffère selon les catégories d'âge pour les deux semestres. Pour déterminer si ces différences observées sont statistiquement significatives, un test ANOVA(Analysis Of Variance) à un facteur a été réalisé.

1. Hypothèses

- H_0 : Les moyennes générales(S1 et S2) sont égales pour tous les groupes d'âge

$$\mu_{18-20} = \mu_{21-23} = \mu_{24-26}$$

- H_1 : Au moins un groupe d'âge a une moyenne différente

2. Seuil de signification: $\alpha = 0.05$

3. Résultat

- p-value S1 = 0.931
- p-value S2 = 0.436

4. Règle de décision

- Si p-value < 0.05 \rightarrow rejet de H_0

Conclusion S2 : Comme p-value(S1 et S2) > 0.05, on ne rejette pas H_0 . Les performances ne diffèrent pas significativement selon les groupes d'âge.

3. Analyse des facteurs de réussite

a) Impact des heures d'étude sur la performance.

Dans cette partie nous allons voir s'il existe un lien (relation linéaire) entre les heures d'étude et la performance académique. Pour faire nous allons utiliser le Coefficient de corrélation de Pearson.

Corrélation entre heures d'étude et performance pour les deux semestres.

Semestre	Coefficient de corrélation (Pearson)	Force	Interprétation
S1	0.375	Modérée	Corrélation positive modérée : les étudiants qui étudient plus tendent à obtenir de meilleures moyennes.
S2	0.393	Modérée	Corrélation positive modérée : les étudiants qui étudient plus tendent à obtenir de meilleures moyennes.

Le tableau montre une relation positive modérée entre le nombre d'heures d'étude hebdomadaires et la performance académique pour les deux semestres.

- Cela signifie que plus un étudiant consacre de temps à l'étude plus sa moyenne tend à être élevée même si l'effet n'est pas extrêmement fort.

- Le coefficient de corrélation reste inférieur à 0.5 ce qui suggère que d'autres facteurs peuvent également influencer la réussite.

b) Impact des absences

Dans cette partie nous allons voir s'il existe un lien (relation linéaire) entre le nombre d'absences et la performance académique. Pour faire nous allons utiliser le Coefficient de corrélation de Pearson.

Corrélation Pearson entre nombre d'absences et moyenne générale

Semestre	Coefficient de corrélation	Force	Interprétation
S1	-0.347	Modérée	Une corrélation modérée négative : plus d'absences → moyenne plus faible
S2	-0.203	Faible	Une corrélation faible négative : plus d'absences → légère baisse de la moyenne

Conclusion : Les absences ont un impact négatif sur la performance académique. L'effet est plus marqué au semestre 1 (corrélation modérée) qu'au semestre 2 (corrélation faible) indiquant que la présence régulière en cours est un facteur important pour maintenir un bon niveau de performance.

6) Recommendations

À partir des analyses statistiques réalisées, plusieurs leviers d'action peuvent être mis en place pour améliorer la performance académique des étudiants. Les recommandations ci-dessous s'appuient directement sur les résultats observés concernant les filières, le genre, les heures d'étude, les absences et les facteurs associés à la réussite.

- Renforcer l'accompagnement académique dans les filières moins performantes
- Encourager une meilleure gestion du temps et augmenter l'implication académique
- Réduire le taux d'absences et renforcer le suivi de présence
- Renforcer les stratégies pédagogiques dans les matières complexes

7) Conclusion : Synthèse et limites de l'étude.

L'étude visait à identifier les facteurs influençant la performance académique des étudiants sur deux semestres. Les résultats montrent que les heures d'étude, la régularité de la présence en cours et la filière suivie jouent un rôle déterminant. Les étudiants consacrant 35 à 40 heures hebdomadaires d'étude obtiennent généralement les meilleurs résultats tandis que les absences ont un effet négatif même si la corrélation demeure faible et modérée. Ces constats soulignent l'importance du suivi pédagogique et du soutien institutionnel pour améliorer la réussite académique.

Cependant l'étude présente des limites : les données reposent sur des déclarations susceptibles d'être biaisées, l'approche quantitative ne capture pas les aspects qualitatifs essentiels (motivation, qualité de

l'enseignement, conditions de vie etc...) et les corrélations ne permettent pas d'établir des relations causales.

Malgré cela, l'étude apporte une compréhension générale pertinente et ouvre la voie à des recherches futures utilisant des méthodes mixtes pour mieux saisir la complexité des facteurs de réussite.