

Rapport d'Analyse de la Performance Étudiante

Réalisé par :

Bendy SERVILUS

Etudiant SDMIA

Langage R

Sous la direction du professeur Dr. John Benson LOUIS

Soumis le : 08 décembre 2025

Résumé

Ce rapport présente l'analyse des performances académiques de 200 étudiants sur deux semestres, en mettant en évidence les facteurs influençant la réussite, les évolutions entre S1 et S2, et des recommandations basées sur des analyses statistiques robustes.

Introduction

2.1 Contexte du Projet

L'université souhaite identifier les facteurs clés influençant la réussite académique. Le présent projet analyse les données de 200 étudiants sur deux semestres consécutifs (S1 et S2) afin d'orienter les stratégies d'amélioration des résultats.

2.2 Objectifs et Questions de Recherche

Ce rapport vise à :

1. Caractériser les profils démographiques et comportementaux.
2. Analyser l'évolution des performances entre S1 et S2.
3. Identifier les facteurs associés à la réussite (heures d'étude, absences).
4. Formuler des recommandations concrètes.

Questions de recherche : Facteurs associés à la performance, différences significatives entre filières, évolution S1→S2, impact des heures d'étude et des absences, profils à risque.

Méthodologie

3.1 Description des Données

Les données proviennent du fichier `etudiants_performance.csv` (dossier `/data/raw/`). Le jeu de données contient 200 observations et 24 variables : démographie, comportement, absences, heures d'étude, notes académiques (5 matières sur S1 et S2).

3.2 Préparation et Nettoyage des Données

1. Importation du fichier CSV.
2. Création de variables clés (`moyenne_s1`, `moyenne_s2`, `variation_moyenne`, réussite S1/S2).
3. Conversion des variables catégorielles et binaires en facteurs.

3.3 Outils et Techniques Statistiques

Analyses effectuées en R avec `dplyr`, `ggplot2`, `tidyr`. Les tests incluent : test T apparié, test T indépendant, ANOVA, Tukey HSD, corrélation Pearson.

Fonctions personnalisées utilisées : `resume_statistique()`, `calcul_taux_reussite()`, `exporter_graphique()`.

Résultats

4.1 Statistiques Descriptives

Résumé de l'âge (mean, median, sd), répartition par genre et filière.

Tableau 1.1 : Résumé de l'âge

A	B	C	D	E	F	G	H	I
n	n_manquants	moyenne	mediane	ecart_type	minimum	maximum	q1	q3
200	0	20.66	20	1.57	18	26	20	21

Tableau 1.2 : Répartition par Genre

File/Url:		
~/Desktop/data_analysis_student/output/tables/repartition_genre.csv		
Data Preview:		
Genre (character) ▾	Effectif (double) ▾	Pourcentage (double) ▾
Femme	111	55.5
Homme	89	44.5

Tableau 1.3 : Répartition par ville origine

File/Url:		
~/Desktop/data_analysis_student/output/tables/repartition_villes.csv		
Data Preview:		
Ville (character) ▾	Effectif (double) ▾	Pourcentage (double) ▾
Nantes	55	27.5
Paris	48	24.0
Bordeaux	34	17.0
Lyon	25	12.5
Toulouse	23	11.5
Marseille	15	7.5

Distribution géographique : Nantes (27.5%), Paris (24.0%), Bordeaux (17.0%).

4.3 Performance par Groupes

Analyse ANOVA (moyenne S2 ~ filière). Résultat : L'ANOVA est [significative / non-significative] avec p-value = _____. La filière la plus performante est _____.

Tableau 3.1 : Statistiques et Taux de Réussite par semestre

File/Url:			
~/Desktop/data_analysis_student/output/tables/taux_reussite.csv			
Data Preview:			
Semestre (character) ▾	Taux_reussite (double) ▾	Reussis (double) ▾	Echoues (double) ▾
S1	86	172	28
S2	86	172	28
Evolution	0	NA	NA

Discussion

Interprétation détaillée des résultats obtenus dans les analyses descriptives, comparatives et corrélatives. Focalisation sur le rôle critique des absences et sur les écarts entre filières.

Recommandations

1. Mettre en place un système d'alerte pour les étudiants dépassant un seuil critique d'absences.
2. Allouer des ressources supplémentaires à la filière la moins performante.
3. Organiser des ateliers sur les méthodes d'étude pour les étudiants ayant un faible rendement malgré un volume élevé d'heures d'étude.

Conclusion

Les analyses démontrent que les absences, les heures d'étude et la filière sont des déterminants significatifs de la performance académique. Une progression $S1 \rightarrow S2$ est observée, avec un rôle majeur joué par l'assiduité.

Annexes

Extrait du script principal utilisé pour les analyses.

```
# SCRIPT PRINCIPAL DU PROJET - PERFORMANCE ETUDIANTE  
source('scripts/00_fonctions.R')  
source('scripts/01_import_nettoyage.R')  
source('scripts/02_analyses_descriptives.R')  
source('scripts/03_analyses_comparatives.R')  
source('scripts/04_visualisations.R')
```