

Mini-Projet en Statistique Descriptive

Mode de vie quotidien et performance académique des étudiants

Réalisé par: Adnane Louardi
Code apogée: 22501067
Email: adnane.louardi-etu@univh2c.ma

¹ FSBM Casablanca - 2024/2025

Abstract. *vue détaillée des modes de vie des étudiants et de leur corrélation avec les résultats scolaires*

1. Introduction

Dans le cadre de ce projet, nous nous proposons d'explorer en profondeur les modes de vie de 2000 étudiants et d'étudier leur lien potentiel avec les résultats scolaires. Cette étude s'inscrit dans un contexte où les facteurs influençant la réussite académique suscitent un intérêt croissant, tant au sein de la communauté éducative que dans la société en général.

Notre objectif principal est d'appliquer les concepts statistiques acquis en cours à une problématique concrète et d'en extraire des informations pertinentes. En analysant les données recueillies, nous chercherons à identifier les corrélations éventuelles entre différents aspects des modes de vie étudiante (heures d'étude, activités extrascolaires, sommeil, socialisation, activité physique, niveau de stress) et les performances académiques, mesurées par la note générale (GPA).

On peut faire ceci avec des logiciel ou avec des langages des programmation, j'ai choisi le langage Python et sa bibliothèque Pandas pour traiter les données, et j'ai utilisé Excel comme outil pour exécuter les calculs. Et voilà la structure de ce rapport:

1. Introduction
2. Analyses et résultats
 - (a) Caractères Statistiques
 - i. Les caractères quantitatifs
 - ii. Les caractères qualitatives
 - (b) Données et Méthodologie
 - i. Source des Données
 - ii. Organisation des Données
 - iii. Outils Utilisés
 - (c) Analyse Statistique
 - i. Analyse des Corrélations
 - ii. Analyse de Régression
 - iii. Distributions Marginales
 - (d) Visualisations
3. Conclusion
4. Références

2. Analyse et Résultats

2.1. Caractères Statistiques

- **Les caractères quantitatifs**
 - Nombre des heures d'étude par jour.
 - Nombre des heures de sommeil par nuit.
 - Nombre des heures de socialisation par jour.
 - Nombre des heures d'activité physique par jour.
 - Nombre des heures Participation à des activités extrascolaires.
 - Note générale (GPA).
- **Les caractères qualitatives**
 - Niveau de stress perçu (Faible, Modéré, Haut).

2.2. Données et Méthodologie

2.2.1. Source des Données:

L'ensemble de données utilisé pour cette étude provient de la plateforme Kaggle. Il s'intitule '**Student Lifestyle Dataset**'. Ce ensemble de données contient des informations détaillées sur les habitudes de vie des étudiants, notamment leurs niveaux de stress, heures d'étude par jour, activités physiques, activités parascolaires, la note générale et sommeil. Il a été choisi pour sa pertinence dans l'analyse des comportements et des corrélations dans le cadre de mon projet.

2.2.2. Organisation des Données:

Les données sont structurées et préparées comme suit:

Algorithme 1. Afficher les premières lignes pour vérifier les données

```
1 import pandas as pd
2 import numpy as np
3 df = pd.read_csv('student_lifestyle_dataset.csv')
4 df.head()
```

	Student_ID	Study_Hours_Per_Day	Extracurricular_Hours_Per_Day	Sleep_Hours_Per_Day	Social_Hours_Per_Day	Physical_Activity_Hours_Per_Day	GPA	Stress_Level
0	1	6.9	3.8	8.7	2.8	1.8	2.99	Moderate
1	2	5.3	3.5	8.0	4.2	3.0	2.75	Low
2	3	5.1	3.9	9.2	1.2	4.6	2.67	Low
3	4	6.5	2.1	7.2	1.7	6.5	2.88	Moderate
4	5	8.1	0.6	6.5	2.2	6.6	3.51	High

Figure 1. Les premières lignes d'ensemble de données

On va classer toutes les caractères en 10 classes de même amplitude.

L'algorithme suivante nous aide à réaliser cette classement et faire des visualisations plus tard :

Algorithme 2. Classement des caractères

```
1 //Nombre des divisions
2 bins=[5,5.5,6,6.5,7,7.5,8,8.5,9,9.5,10]
3
4 //Les classes
5 labels=['[5;5.5[',...,'[9.5;10]']
6
7 df['Study_Hours_Per_Day']=pd
8 .cut(df['Study_Hours_Per_Day'], bins=bins, labels=labels)
9
10 print(df['Study_Hours_Per_Day'])
11 .value_counts().sort_index()
```

- Nombre des heures d'étude par jour (selon **Algorithme 2.**):

```
Study_Hours_Per_Day
[5;5.5[      197
[5.5;6[      192
[6;6.5[      213
[6.5;7[      210
[7;7.5[      222
[7.5;8[      180
[8;8.5[      198
[8.5;9[      213
[9;9.5[      181
[9.5;10]     174
Name: count, dtype: int64
```

Figure 2. Classement des heures d'étude par jour

- Nombre des heures de sommeil par nuit:

```
Sleep_Hours_Per_Day
[5.0;5.5[     203
[5.5;6.0[     213
[6.0;6.5[     192
[6.5;7.0[     183
[7.0;7.5[     188
[7.5;8.0[     216
[8.0;8.5[     213
[8.5;9.0[     183
[9.0;9.5[     187
[9.5;10.0]    199
Name: count, dtype: int64
```

Figure 3. Classement des heures de sommeil par nuit

- Nombre des heures de socialisation par jour:

```

Social_Hours_Per_Day
[0.0;0.6[      256
[0.6;1.2[      232
[1.2;1.8[      229
[1.8;2.4[      214
[2.4;3.0[      208
[3.0;3.6[      209
[3.6;4.2[      180
[4.2;4.8[      183
[4.8;5.4[      151
[5.4;6.0]      120
Name: count, dtype: int64

```

Figure 4. Classement des heures de socialisation par jour

- Nombre des heures d'activité physique par jour:

```

Physical_Activity_Hours_Per_Day
[0.0;1.3[      247
[1.3;2.6[      331
[2.6;3.9[      360
[3.9;5.2[      370
[5.2;6.5[      285
[6.5;7.8[      209
[7.8;9.1[      117
[9.1;10.4[      56
[10.4;11.7[      15
[11.7;13.0]      4
Name: count, dtype: int64

```

Figure 5. Classement des heures d'activité physique par jour

- Nombre des heures de participation à des activités extrascolaires:

```

Extracurricular_Hours_Per_Day
[0.0;0.4[      216
[0.4;0.8[      194
[0.8;1.2[      176
[1.2;1.6[      226
[1.6;2.0[      198
[2.0;2.4[      204
[2.4;2.8[      196
[2.8;3.2[      183
[3.2;3.6[      220
[3.6;4.0]      161
Name: count, dtype: int64

```

Figure 6. Classement des heures de participation à des activités extrascolaires

- Note générale (GPA):

GPA	
[2.24;2.42[11
[2.42;2.59[60
[2.59;2.77[182
[2.77;2.94[347
[2.94;3.12[440
[3.12;3.30[420
[3.30;3.47[290
[3.47;3.65[179
[3.65;3.82[59
[3.82;4.00[11
Name: count, dtype: int64	

Figure 7. Classement des Notes générales (GPA)

- Niveau de stress:

Stress_Level	
High	1029
Moderate	674
Low	297
Name: count, dtype: int64	

Figure 8. Classement des niveaux de stress

2.2.3. Outils Utilisés:

- Kaggle: Utilisé pour obtenir l'ensemble de données.
- Excel: Utilisé pour effectuer les calculs statistiques (corrélations, régressions) et créer des graphiques illustratifs.
- Python (Pandas): Utilisé pour classement et vérification des valeurs manquantes dans l'ensemble de données.
- Latex: Utilisé pour la rédaction et la mise en forme de ce rapport.

2.3. Analyse Statistique:

2.3.1. Analyse de Corrélation:

Caractère 1	Caractère 2	Coefficient de Corrélation (r)	Interprétation
Heures d'étude	Activités extrascolaires	-0,002628948	Corrélation faible
Heures d'étude	Sommeil	0,026717287	Corrélation faible
Heures d'étude	Socialisation	-0,137820443	Corrélation faible
Heures d'étude	Activité physique	-0,488113277	Corrélation négative modérée
Heures d'étude	Note générale	0,734467981	Corrélation positive modérée
Activités extrascolaires	Sommeil	0,00884428	Corrélation faible
Activités extrascolaires	Socialisation	-0,139080746	Corrélation faible
Activités extrascolaires	Activité physique	-0,369988799	Corrélation faible
Activités extrascolaires	Note générale	-0,032173532	Corrélation faible
Sommeil	Socialisation	-0,193555696	Corrélation faible
Sommeil	Activité physique	-0,470302459	Corrélation négative modérée
Sommeil	Note générale	-0,004278441	Corrélation faible
Socialisation	Note générale	-0,08567714	Corrélation faible
Socialisation	Activité physique	-0,417142088	Corrélation négative modérée
Activité physique	Note générale	-0,341152464	Corrélation faible

J'ai calculé les coefficients de corrélation pour toutes les paires de caractères. Les deux résultats les plus significatifs sont présentés ci-dessous:

Caractère 1	Caractère 2	Coefficient de Corrélation (r)	Interprétation
Heures d'étude	Note générale	0,734467981	Corrélation positive modérée
Heures d'étude	Activité physique	-0,488113277	Corrélation négative modérée

1. Corrélation entre les Heures d'Étude et la Note Générale (GPA):
Une corrélation forte indique que plus les heures d'étude augmentent, plus la moyenne générale a tendance à être élevée.
2. Corrélation entre les Heures d'Étude et les Activités Physiques:
Cela suggère que plus les heures d'étude augmentent, moins de temps est consacré aux activités physiques.

Les résultats final sont collectée dans le tableau suivant:

	Heures d'étude par jour	Heures parascolaires par jour	Heures de sommeil par jour	Heures sociales par jour	Heures d'activité physique par jour	GPA
Heures d'étude par jour	1					
Heures parascolaires par jour	-0,002628948	1				
Heures de sommeil par jour	0,026717287	0,00884428	1			
Heures sociales par jour	-0,137820443	-0,139080746	-0,193555696	1		
Heures d'activité physique par jour	-0,488113277	-0,369988799	-0,470302459	-0,417142088	1	
GPA	0,734467981	-0,032173532	-0,004278441	-0,08567714	-0,341152464	1

Figure 9. Les calculs de corrélation pour chaque paire de caractères

2.3.2. Analyse de Régression:

Concernant cette section, on vas déterminer les équations de régressions pour la paire de caractères le plus corrélés.

1. Pour la paire "heures d'étude" (X) et "moyenne générale" (Y), l'équation de régression est donnée par :

$$Y = 1,92 + 0.15X$$

Exemple:

En utilisant l'équation, un étudiant a étudié pendant 8 heures par jour, la moyenne générale prévue est:

$$Y = 1,92 + 0.15 * 8 = 3,19$$

Cela suggère qu'une augmentation significative des heures d'étude peut conduire à une amélioration notable des résultats académiques.

2. Pour la paire "activités physique" (X) et "heures d'étude" (Y), l'équation de régression est donnée par :

$$Y = 8,67 - 0.27X$$

Exemple:

Un étudiant a été active physiquement pendant 4 heures par jour, les heures d'étude prévue sont:

$$Y = 8,67 - 0.27 * 4 = 7,5$$

Cela montre une relation inverse entre le temps d'étude et l'engagement physique.

2.3.3. Distributions Marginales:

Dans cette section, nous analysons les distributions marginales des deux caractères les plus corrélés, à savoir les heures d'étude et la moyenne générale (GPA). Pour chaque caractère, nous calculons les principaux paramètres de position, de dispersion et de forme afin d'obtenir une compréhension approfondie de leurs comportements statistiques.

1. Paramètres de Position:

Les paramètres de position, tels que la moyenne et la médiane, permettent d'identifier les valeurs centrales des distributions. Voici les résultats pour les deux caractères, calculé à l'aide des fonctions Excel **MOYENNE()** et **MÉDIANE()** :

- Moyenne:
Pour Heures d'Étude: 7,47
Pour Moyenne Générale (GPA): 3,11
- Médiane:
Pour Heures d'Étude: 7,40
Pour Moyenne Générale (GPA): 3,11

Interprétation générale: Les moyennes et les médianes proches montrent une répartition équilibrée des GPA et des heures d'étude. Cela reflète une cohérence générale dans les performances académiques et les habitudes d'étude des étudiants.

2. Paramètres de Dispersion:

Les paramètres de dispersion, comme la variance et l'écart-type, mesurent la variabilité au sein des données, calculé à l'aide des fonctions Excel **VAR.P()** et **ECARTYPEP()** :

- Variance:
Pour Heures d'Étude: 2,02
Pour Moyenne Générale (GPA): 0,08
- Écart-type:
Pour Heures d'Étude: 1,42
Pour Moyenne Générale (GPA): 0,29

Interprétation générale: L'écart-type plus faible pour les GPA indique une homogénéité dans les performances académiques, tandis que l'écart-type plus élevé pour les heures d'étude traduit une diversité plus marquée dans les pratiques d'étude des étudiants.

3. Paramètres de Forme:

Les paramètres de forme, notamment l'asymétrie et l'aplatissement, donnent des informations sur la symétrie et la concentration de la distribution, calculé à l'aide des fonctions Excel **COEFFICIENT.ASYMETRIE()** et **KURTOSIS()** :

- Coefficient d'asymétrie:
Pour Heures d'Étude: 0,03
Pour Moyenne Générale (GPA): 0,02
- Coefficient d'aplatissement:
Pour Heures d'Étude: -1,18
Pour Moyenne Générale (GPA): -0,37

Interprétation générale: Les asymétries faibles et les distributions légèrement aplaties montrent qu'il n'y a pas de valeurs extrêmes significatives pour les GPA ou les heures d'étude, ce qui confirme une répartition équilibrée et régulière des données.

Les résultats final sont collectée dans le tableau suivant:

	les distributions marginales	GPA	Heures d'étude par jour	La fonction utilisée
Paramètre de position	moyenne	3,11596	7,4758	MOYENNE()
	médiane	3,11	7,4	MEDIANE()
Paramètre de dispersion	variance	0,089161278	2,02644436	VAR.P()
	écart-type	0,298598859	1,423532353	ECARTYPEP()
Paramètre de forme	coefficients d'asymétrie	0,027515856	0,034987842	COEFFICIENT.ASYMETRIE()
	coefficients d'aplatissement	-0,37068704	-1,181323271	KURTOSIS()

Figure 10. Analyse des distributions marginales

2.4. Visualisations:

Cette section présente les graphiques utilisés pour analyser et visualiser les relations entre les différentes variables étudiées, ainsi que les distributions marginales. Ces visualisations permettent de mieux comprendre les tendances et les caractéristiques des données.

- Nuages de points:

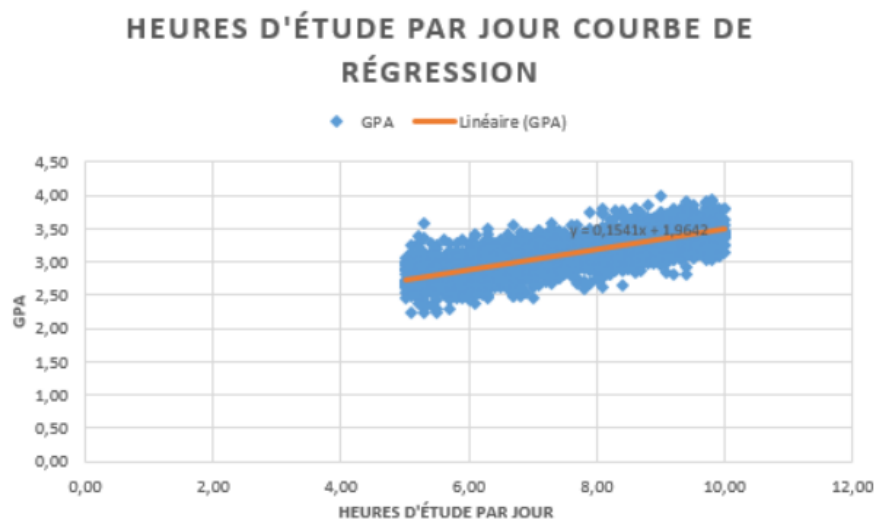


Figure 11. Relation entre les heures d'étude par jour et le GPA.

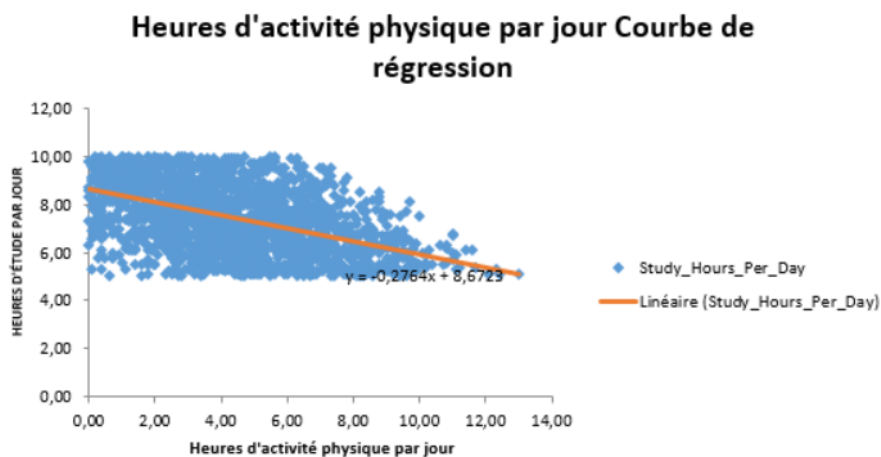


Figure 12. Relation entre les heures d'étude et les activités physiques.

- Histogrammes :

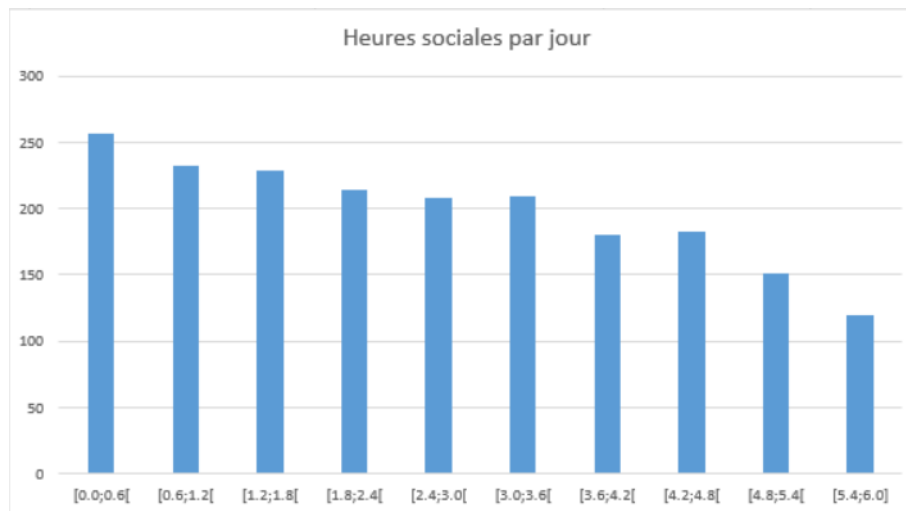


Figure 13. La répartition des heures social quotidienne.

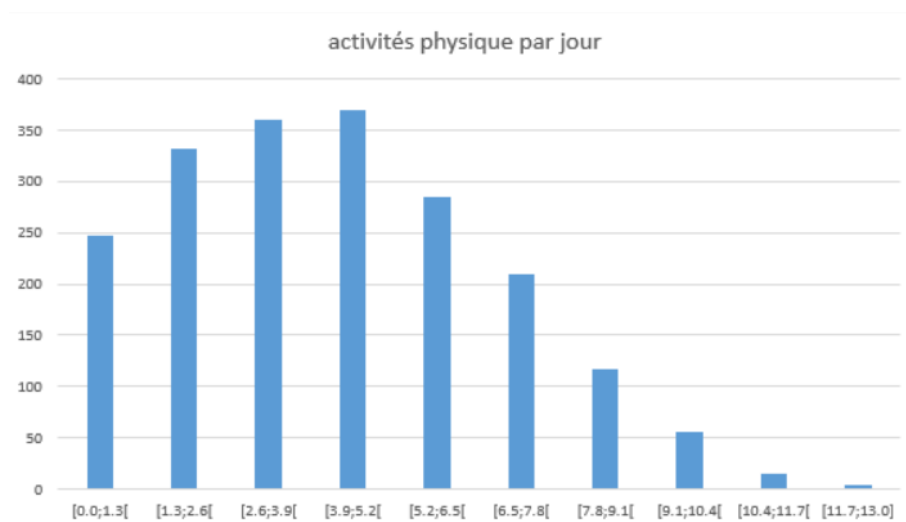


Figure 14. La répartition des heures d'activité physique.

- Graphique en secteur :

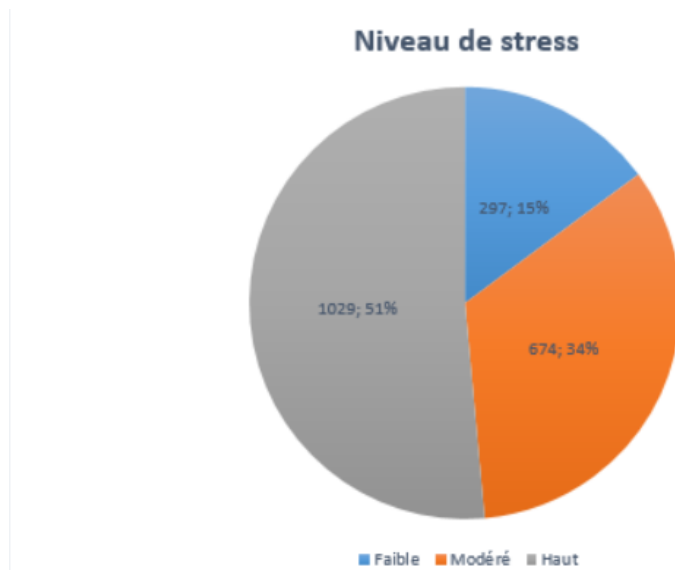


Figure 15. La répartition du niveau de stress.

- Boîte à moustaches :

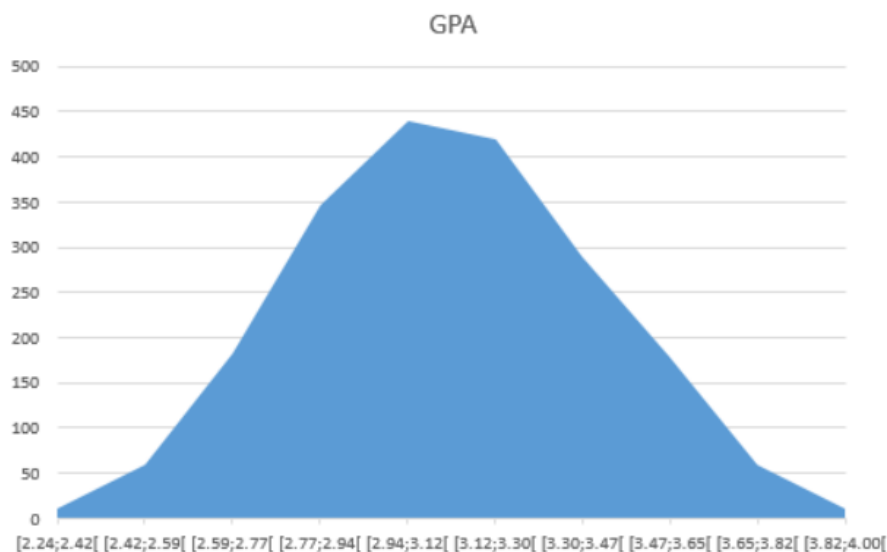


Figure 16. La répartition de la note générale (GPA) .

3. Conclusion

Ce projet met en évidence l'importance des habitudes d'étude et de vie équilibrées pour garantir de bonnes performances académiques. Il illustre également comment des outils simples, mais puissants, comme Python, Excel et LaTeX, peuvent être utilisés pour mener des analyses approfondies et produire des documents professionnels. En combinant des analyses statistiques rigoureuses avec des visualisations claires, ce travail fournit une base solide pour mieux comprendre les dynamiques étudiantes et suggère des pistes d'amélioration pour le bien-être et la réussite des étudiants.

4. References

- Kaggle comme un source de donnees: <https://www.kaggle.com/code/steve1215rogg/data-analysis-on-student-s-daily-lifestyle>
- LaTeX pour la rédaction du document: <https://www.latex-project.org/> et <https://www.overleaf.com>
- Pandas pour la traitement de Data: <https://pandas.pydata.org/>