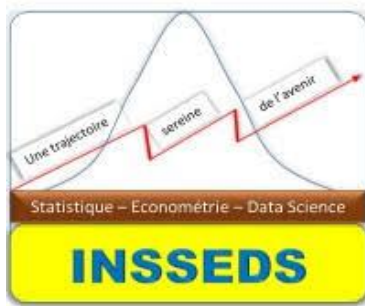


**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE**

REPUBLIQUE DE CÔTE D'IVOIRE



**INSTITUT SUPERIEUR DE
STATISTIQUE D'ECONOMETRIE
ET DATA SCIENCE**



UNION-DISCIPLINE-TRAVIL

MASTER 1

INGENIEUR STATISTICIEN-DATA ANALYST

MINI-PROJET

STATISTIQUE DESCRIPTIVE : UNIVARIEE ET BIVARIEE

**ETUDUE DES FACTEURS QUI
INFLUENCENT LA DEPRESSION CHEZ LES**

**ANNEE ACADEMIQUE
2024-2025**

**NOM : DAGNOGO
PRENOM : Clagniri Safiatou**

**Enseignant- Encadreur
AKPOSSO DIDIER MARTIAL**

Avant-Propos

L'analyse des facteurs influençant la dépression chez les étudiants est devenue un enjeu crucial pour mieux comprendre les causes et les mécanismes de cette problématique de santé mentale. Dans un contexte où les pressions académiques, sociales et économiques pèsent de plus en plus sur les étudiants, il est essentiel d'identifier les déterminants majeurs de la dépression afin de proposer des stratégies de prévention et d'accompagnement adaptées. Une analyse statistique rigoureuse permet d'examiner les liens entre divers facteurs, tels que les habitudes de vie, les antécédents familiaux ou encore le niveau de stress, afin d'identifier les profils les plus à risque.

Cette étude vise à analyser en profondeur les éléments contribuant à la dépression chez les étudiants en s'appuyant sur un ensemble de données détaillées. En exploitant des méthodes statistiques avancées, l'objectif est d'identifier les tendances et les corrélations entre les différents facteurs de risque. Cette approche permettra de mieux cerner les groupes d'étudiants les plus vulnérables et d'orienter les actions de sensibilisation et d'intervention de manière plus efficace. L'analyse mettra en évidence les variables les plus significatives dans l'apparition de la dépression et offrira des pistes pour développer des stratégies de prévention adaptées aux besoins spécifiques des étudiants.

Tables des matières

Avant-propos.....	2
INTRODUCTION.....	5
I) PREPARATION DES DONNEES.....	7
1) Visualisation des valeurs manquantes.....	7
2) Traitement des valeurs manquantes.....	7
3) Traitement des valeurs aberrantes où extrêmes.....	8
II) ANALYSE EXPLORATOIRE DES DONNEES.....	8
A) Analyse Univariée.....	9
1) La variable âge.....	9
1.1) Tableau statistique.....	9
1.2) Graphique.....	10
1.3) Résumé numérique et interprétation.....	12
2) La variable nombre d'heure de travail/études.....	12
2.1) Tableau statistique.....	12
2.2) Graphique.....	13
2.3) Résumé numérique et interprétation.....	14
3) La variable sexe.....	15
4) La variable pression académique.....	15
5) La variable stress financier.....	16
6) La variable pensée suicidaire.....	16
7) La variable habitudes alimentaires.....	16
B) Analyse Bivariée.....	17
1) Est-ce que la dépression chez les étudiants est liée à l'âge ?.....	17
2) Est-ce que la dépression chez les étudiants dépend du sexe ?.....	18
3) Est-ce que la dépression chez les étudiants dépend de la pression académique ?.....	18
4) Est-ce que la dépression chez les étudiants est liée au nombre d'heure de travail/études ?..	19
5) Est-ce que la dépression chez les étudiants est causée par la durée de sommeil ?.....	19
6) Est-ce que la dépression chez les étudiants est liée au stress financier ?.....	20
7) Est-ce que la dépression chez les étudiants dépend de l'habitude alimentaire ?.....	20
8) Est-ce que la dépression chez les étudiants est liée aux antécédants familiaux de maladie mentale ?.....	21
9) Est-ce que la dépression chez les étudiants est liée à la satisfaction ?.....	21
CONCLUSION.....	22
ANNEXE.....	23

Liste des tableaux

Tableau 1 : Extrait du jeu de donnée.

Tableau 2 : Tableau des effectifs et fréquences pour la variable âge.

Tableau 3 : Paramètre statistique pour la variable âge.

Tableau 4 : Tableau des effectifs et fréquences pour la variable nombre d'heure de travail/études.

Tableau 5 : Paramètre statistique pour la variable nombre d'heure de travail/études.

Tableau 6 : Tableau de contingence de la variable dépression et âge.

Liste des figures

Figure 1 Visualisation des valeurs manquantes

Figure 2 Traitement des valeurs manquantes

Figure 3 Traitement des valeurs extrêmes

Figure 4 Traitement des valeurs extrêmes

Figure 3 Student_depression\$age

Figure 4 Nombre_heure_travail_etude

Figure 5 Visualisation de la distribution du sexe

Figure 6 Visualisation de la distribution de la pression académique

Figure 7 Visualisation de la distribution du stress financier

Figure 8 Distribution de la pensée suicidaire

Figure 9 Distribution de l'habitude alimentaire

Figure 10 Boîte à moustache par niveau de facteur

Figure 11 Diagramme en bâtons groupés dépression et sexe

Figure 12 Diagramme en bâtons groupés dépression et pression académique

Figure 13 Diagramme en bâtons groupés nombre_heure_travail_etude

Figure 14 Diagramme en bâtons groupés dépression et durée de sommeil

Figure 15 Diagramme en bâtons groupés dépression et stress financier

Figure 16 Diagramme en bâtons groupés dépression et habitude alimentaire

Figure 17 Diagramme en bâtons groupés dépression et antécédants familiaux maladie mentale

Figure 18 Diagramme en bâtons groupés dépression et satisfaction études

INTRODUCTION

Dans un contexte où la santé mentale des étudiants est devenue une préoccupation majeure, il est essentiel d'identifier les facteurs influençant la dépression afin de mieux comprendre ses causes et ses mécanismes. L'analyse statistique des données permet d'examiner les différentes variables susceptibles d'expliquer l'apparition et l'évolution des troubles dépressifs chez les étudiants. En prenant en compte des critères tels que les conditions de vie, les habitudes de sommeil, les antécédents familiaux ou encore le niveau de stress, cette étude vise à dégager des tendances significatives pour mieux prévenir et traiter la dépression.

Face à la diversité des expériences et des facteurs de risque associés à la dépression, la principale problématique de cette étude est de déterminer quels sont les éléments les plus influents et comment ils interagissent. L'objectif est d'identifier des groupes d'étudiants partageant des caractéristiques similaires en termes de vulnérabilité psychologique et d'évaluer les corrélations entre les différentes variables étudiées. Une telle approche permettra d'adapter les stratégies de prévention et d'intervention pour mieux répondre aux besoins des étudiants les plus exposés.

Les résultats attendus de cette étude sont :

- Une classification détaillée des étudiants selon leur vulnérabilité face à la dépression.
- L'identification des facteurs les plus déterminants dans le développement de troubles dépressifs.
- Des recommandations concrètes pour renforcer les actions de prévention et optimiser l'accompagnement des étudiants en difficulté.

L'étude s'appuie sur une méthodologie rigoureuse, intégrant plusieurs étapes analytiques :

- **Prétraitement des données** : Nettoyage et structuration des informations recueillies.
- **Analyse exploratoire des données (EAD)** : Identification des tendances et des premiers indicateurs significatifs.
- **Segmentation des étudiants** : Classification des individus en fonction de leurs niveaux de risque et de leurs caractéristiques communes.
- **Évaluation des résultats** : Vérification des corrélations et validation des segments identifiés.

Dictionnaire des données

Un dictionnaire de données est un document qui présente une description complète de chaque variable utilisée dans une analyse statistique ou économétrique. Il détaille les propriétés, les caractéristiques et le contexte de chaque variable, ainsi que leur signification.

Variable	Description	Type
Id	Identifiant unique de chaque étudiant	Numérique
Sexe	Sexe (Masculin/Féminin)	Catégorielle
Age	Âge de l'étudiant	Numérique
Ville	Ville de résidence	Catégorielle
Profession	Profession (étudiant à temps plein ou autre)	Catégorielle
pression_academique	Pression académique (échelle)	Ordinale / Numérique
pression_liee_au_travail	Pression liée au travail (échelle)	Ordinale / Numérique
moyenne_notes	Moyenne générale des notes	Numérique
satisfaction_etudes	Satisfaction par rapport aux études (échelle)	Ordinale
satisfaction_travail	Satisfaction par rapport au travail (échelle)	Ordinale
duree_sommeil	Durée du sommeil (catégorielle)	Catégorielle Ordinale
habitudes_alimentaires	Habitudes alimentaires (saines, modérées)	Catégorielle Ordinale
diplome_suivi	Diplôme suivi ou obtenu (BSc, M.Tech, etc.)	Catégorielle
pensees_suicidaires?	Pensées suicidaires (Oui/Non)	Binaire
nombre_heure_travail_etude	Nombre d'heures de travail ou d'études par jour	Numérique
stress_financier	Stress financier (échelle ou score)	Ordinale / Numérique
antecedents_familiaux_maladies_mentales	Antécédents familiaux de maladies mentales (Oui/Non)	Binaire
Depression	Dépression (1 pour oui, 0 pour non)	Binaire

Tableau 1 : Extrait du jeu de donnée.

I) PREPARATION DES DONNEES

La préparation des données est une étape essentielle garantissant la qualité et la fiabilité des analyses. Elle comprend la collecte, le nettoyage, la transformation et la structuration des données afin d'éliminer les erreurs, normaliser les formats et sélectionner les informations pertinentes. Un bon prétraitement réduit les biais, améliore

la précision des analyses et optimise les performances des modèles en statistiques. En assurant la cohérence et l'exploitabilité des données, il permet d'obtenir des résultats fiables et d'appuyer des décisions plus éclairées.

1) Visualisation des données manquantes

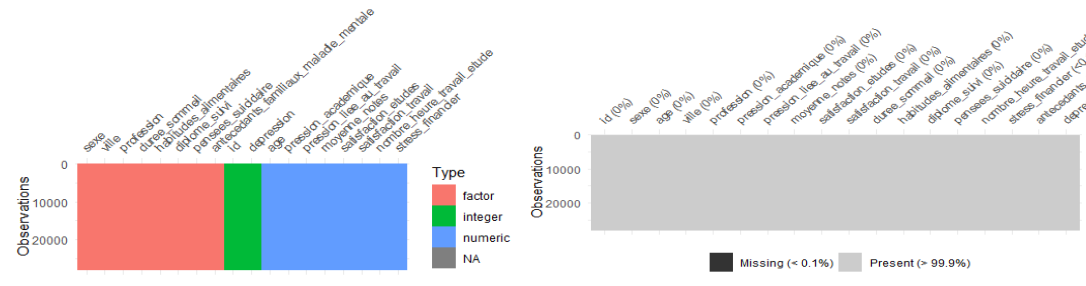


Figure 19 Visualisation des valeurs manquantes

Nous remarquons qu'il y a 3 valeurs manquantes dans le jeu de données plus précisément dans la colonne `stress_financier`, cependant il n'y a pas de doublons. Afin de préserver la distribution d'origine des données, nous allons supprimer les lignes avec des valeurs manquantes de la colonne `stress_financier`.

2) Traitement des valeurs manquantes

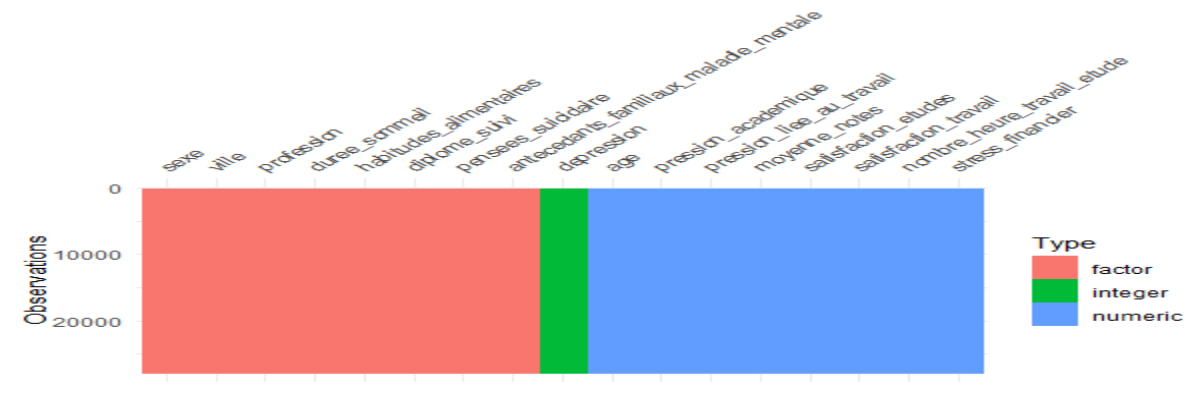


Figure 20 Traitement des valeurs manquantes

3) Traitement des valeurs aberrantes ou extrêmes

Les valeurs aberrantes sont des observations qui diffèrent significativement des autres données dans un jeu de données. Elles révèlent des phénomènes inhabituels ou des cas exceptionnels. Leur identification et traitement est

cruciale pour une bonne analyse car elles peuvent fausser les analyses statistiques, telles que les moyennes et les écarts-types.

Visualisation des valeurs aberrantes

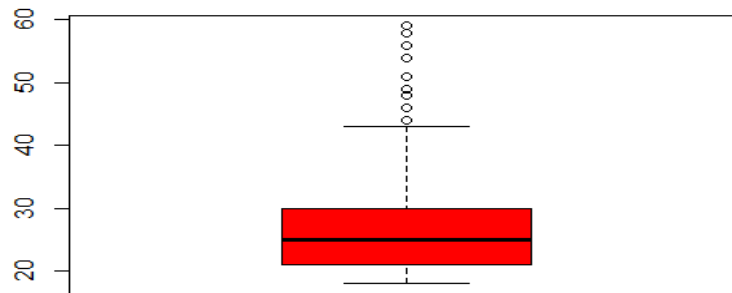


Figure 21 Visualisation des valeurs extrêmes

Traitement

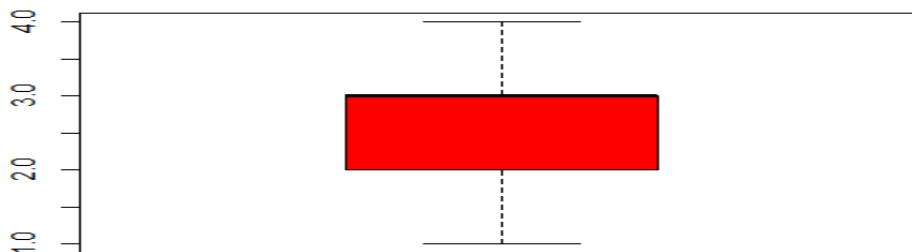


Figure 22 Traitement des valeurs extrêmes

II) ANALYSE EXPLORATOIRE DES DONNEES

Cette partie nous permettra d'avoir certaines informations sur chaque variable quantitative et aussi de déterminer les tendances générales de ces variables.

A) Analyse univariée

1) Analyse de la variable âge

1.1) Tableau statistiqueTableau des effectifs et fréquences

Âge	Effectifs	Eff_Cum_Decrois	Fréquence	Freq_Cum_Crois	Freq_Cum_Decrois
18	1 587	27 898 0.0569		0.0569	1.0000
19	1 560	27 897 0.0559		0.1128	1.0000
20	2 236	27 896 0.0801		0.1930	0.9999
21	1 726	27 895 0.0619		0.2548	0.9999
22	1 160	27 894 0.0416		0.2964	0.9999
23	1 645	27 893 0.0590		0.3554	0.9998
24	2 258	27 892 0.0809		0.4363	0.9998
25	1 784	27 889 0.0639		0.5003	0.9997
26	1 155	27 887 0.0414		0.5417	0.9996
27	1 462	27 886 0.0524		0.5941	0.9996
28	2 133	27 884 0.0765		0.6705	0.9995
29	1 949	27 880 0.0699		0.7404	0.9994
30	1 145	27 879 0.0410		0.7814	0.9993
31	1 427	27 876 0.0512		0.8326	0.9992
32	1 261	27 868 0.0452		0.8778	0.9989
33	1 893	27 866 0.0679		0.9456	0.9989
34	1 468	27 859 0.0526		0.9982	0.9986
35	10	27 849 0.0004		0.9986	0.9982
36	7	26 381 0.0003		0.9989	0.9456
37	2	24 488 0.0001		0.9989	0.8778
38	8	23 227 0.0003		0.9992	0.8326
39	3	21 800 0.0001		0.9993	0.7814
41	1	20 655 0.0000		0.9994	0.7404
42	4	18 706 0.0001		0.9995	0.6705
43	2	16 573 0.0001		0.9996	0.5941
44	1	15 111 0.0000		0.9996	0.5417
46	2	13 956 0.0001		0.9997	0.5003
48	3	12 172 0.0001		0.9998	0.4363
49	1	9 914 0.0000		0.9998	0.3554
51	1	8 269 0.0000		0.9999	0.2964
54	1	7 109 0.0000		0.9999	0.2548
56	1	5 383 0.0000		0.9999	0.1930
58	1	3 147 0.0000		1.0000	0.1128
59	1	1 587 0.0000		1.0000	0.0569

Tableau 2 : Tableau des effectifs et fréquences pour la variable âge

50% des étudiants ont 25 ans ou moins (Freq_Cum_Crois = 0.5002 pour 25 ans).

Plus de 78% des étudiants ont 30 ans ou moins, ce qui confirme que la majorité de la population étudiée est jeune.

Moins de 6% des étudiants ont 34 ans, ce qui signifie que les étudiants plus âgés sont rares dans cet échantillon.

1.2) Graphiques

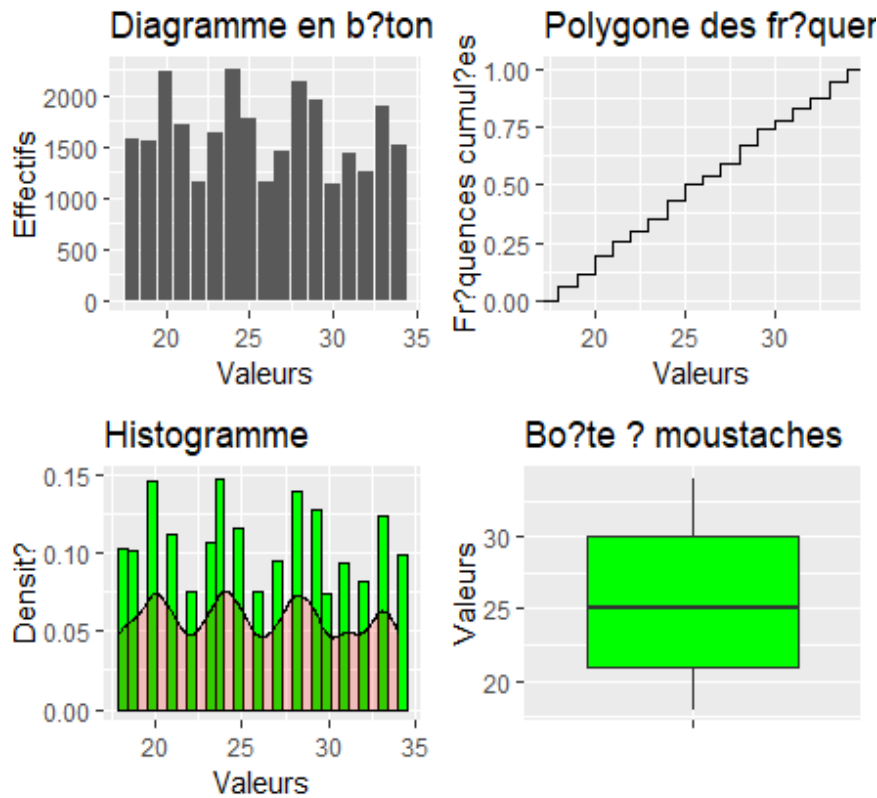


Figure 23 Student_depression\$age

Ces graphiques illustrent la répartition de l'âge des étudiants à travers différentes représentations statistiques, permettant ainsi d'analyser leur distribution et leur dispersion.

Tout d'abord, le **diagramme en bâtons** montre le nombre d'étudiants pour chaque tranche d'âge. On constate une répartition relativement équilibrée, sans qu'un âge précis ne domine nettement. Cela suggère une certaine diversité des âges au sein de l'échantillon.

Ensuite, le **polygone des fréquences cumulées** permet de visualiser la répartition progressive des âges. Il montre une montée régulière des fréquences cumulées, ce qui indique une distribution assez homogène sans ruptures brusques. Cela facilite l'estimation des seuils de répartition et des quartiles.

Par ailleurs, l'**histogramme** complète l'analyse en représentant la densité des âges. On y observe plusieurs pics, suggérant une distribution possiblement multimodale. Cela pourrait signifier la présence de plusieurs groupes d'étudiants, peut-être des jeunes entrants et des étudiants en reprise d'études.

Enfin, la **boîte à moustaches** permet d'évaluer la dispersion des âges. La médiane semble se situer autour de 27-28 ans, tandis que la majorité des étudiants ont entre 20 et 35 ans.

Nous ferons un tableau statistique concernant l'âge des étudiants.

1.3) Resumés numériques et interprétation

Statistique	Valeur
Minimum	18
Maximum	34
Mode (valeur la plus fréquente)	24
Médiane (Q2, 50%)	25
Moyenne	25.81
Quantiles (0%, 25%, 50%, 75%, 100%)	16 - 21 - 25 - 30 - 34
Coefficient de variation (%)	24.71
Variance	4.87
Écart-type	0.066
Coefficient d'asymétrie (Skewness)	18.09 (distribution légèrement étalée à droite)
Coefficient d'aplatissement (Kurtosis)	1.815 (distribution platicurtique)

Tableau 3 : Paramètre statistique pour la variable âge

- **Âge moyen** : 25.81 ans, très proche de la médiane (25 ans), indiquant une distribution assez symétrique.
- **Mode** : 24 ans est l'âge le plus fréquent dans la population étudiée.
- **Écart-type** : 4.87 ans → la majorité des étudiants ont un âge proche de 25 ans, avec une variabilité modérée.
- **Asymétrie (Skewness = 0.066)** :
→ La distribution est légèrement étalée à droite, signifiant qu'il y a quelques étudiants plus âgés qui tirent la moyenne vers le haut.
- **Aplatissement (Kurtosis = 1.815)** :
→ La distribution est platicurtique, donc moins concentrée autour de la moyenne et plus étalée qu'une distribution normale.

2) La variable nombre heure travail etude2.1) Tableau statistiqueTableau des effectifs et fréquences

Heures	Effectifs	Eff_Cum_Crois	Eff_Cum_Decrois	Fréquence	Freq_Cum_Crois	Freq_Cum_Decrois
0	1 699	1 699	27 891	0.0609	0.0609	1.0000
1	1 150	2 849	24 722	0.0412	0.1021	0.8864
2	1 589	4 438	21 830	0.0570	0.1591	0.7827
3	1 469	5 907	17 596	0.0527	0.2118	0.6309
4	1 613	7 520	15 572	0.0578	0.2696	0.5583

5	1 296	8 816	13 064	0.0465	0.3161	0.4684
6	2 247	11 063	11 063	0.0806	0.3967	0.3967
7	2 001	13 064	8 816	0.0717	0.4684	0.3161
8	2 508	15 572	7 520	0.0899	0.5583	0.2696
9	2 024	17 596	5 907	0.0726	0.6309	0.2118
10	4 234	21 830	4 438	0.1518	0.7827	0.1591
11	2 892	24 722	2 849	0.1037	0.8864	0.1021
12	3 169	27 891	1 699	0.1136	1.0000	0.0609

Tableau 4 : Tableau des effectifs et fréquences pour la variable nombre d'heure de travail/études.

- La majorité des étudiants ont entre 6 et 10 heures de travail/études par jour.
- Seuls 6% des étudiants n'ont aucune charge de travail/étude (0 heure).
- 78% des étudiants au moins 10 heures par jour, ce qui indique une forte pression académique.

2.2) Graphiques

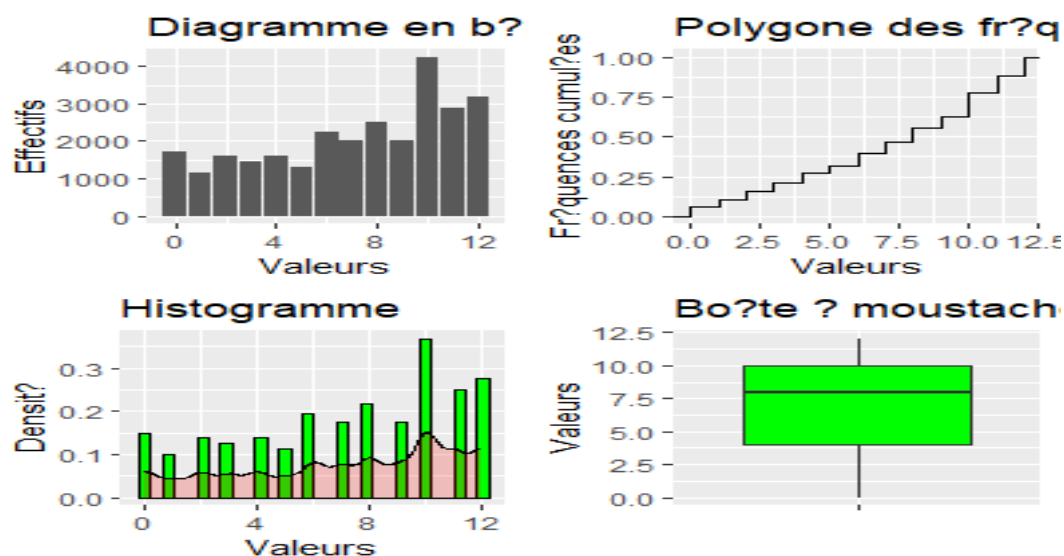


Figure 24 Nombre_heure_travail_etude

On voit le **Diagramme en barres** la répartition des valeurs sous forme de barres. Certaines valeurs reviennent plus souvent que d'autres, notamment vers la fin où le nombre est plus élevé.

La courbe des fréquences cumulées montre l'accumulation des valeurs au fur et à mesure. On voit que ça augmente régulièrement, ce qui veut dire que les valeurs se répartissent de manière assez continue.

Sur l'**Histogramme** on observe la densité des valeurs sur l'intervalle. Certaines plages de valeurs sont plus fréquentes, en particulier vers la droite du graphique. La ligne noire permet de mieux visualiser la tendance générale.

Le dernier graphique qui est la **boîte à moustache** montre la répartition générale des valeurs. La boîte verte représente la majorité des données (50 % du milieu), tandis que les "moustaches" montrent les valeurs extrêmes. Ici, les valeurs semblent plutôt concentrées au centre avec quelques écarts en haut et en bas.

2.3) Résumés numériques et interprétation

Statistique	Valeur
Minimum	0
Maximum	12
Mode (valeur la plus fréquente)	10
Médiane (Q2, 50%)	8
Moyenne	7.16
Quantiles (0%, 25%, 50%, 75%, 100%)	0 - 4 - 8 - 10 - 12
Coefficient de variation (%)	51.80%
Variance	13.74
Écart-type	3.71
Coefficient d'asymétrie (Skewness)	-0.4548 (distribution étalée à gauche)
Coefficient d'aplatissement (Kurtosis)	2.0005 (distribution platicurtique)

Tableau 5 : Paramètre statistique pour la variable nombre d'heure de travail/études.

La moyenne est de 7,16 heures, proche de la médiane (8 heures), ce qui signifie que la distribution est plutôt équilibrée.

Le mode est 10 heures, ce qui signifie que c'est la valeur la plus courante chez les étudiants.

L'écart-type est de 3,71 heures, indiquant une dispersion modérée des heures de travail.

Le coefficient de variation de 51,8% montre une grande variabilité dans les habitudes d'étude des étudiants.

Asymétrie négative (Skewness = -0.4548) → La distribution est étalée à gauche, ce qui signifie que certains étudiants travaillent peu (proche de 0 heures), mais la majorité travaille autour de 8 à 10 heures.

Aplatissement (Kurtosis = 2.0005, platicurtique) → La distribution est moins concentrée autour de la moyenne, indiquant une diversité importante dans les durées d'étude

3) La variable sexe

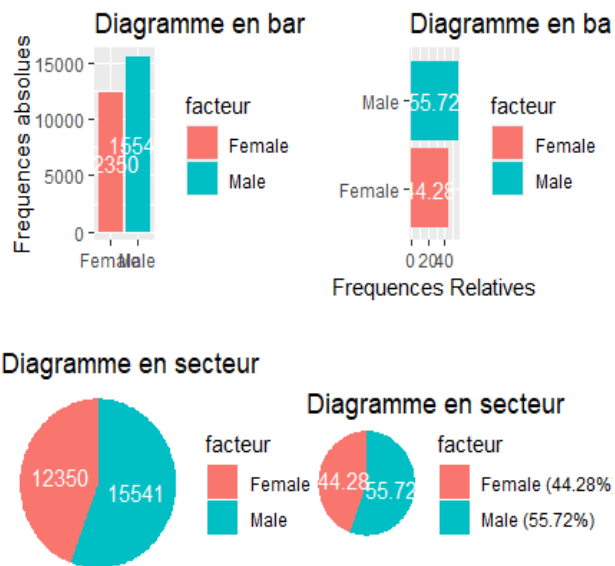


Figure 25 Visualisation de la distribution du sexe

4) La variable pression académique

Environ un tiers (32.2%) des étudiants ressentent une faible pression académique. Cela indique que ces étudiants parviennent peut-être à gérer efficacement leurs responsabilités académiques ou qu'ils ne perçoivent pas les exigences scolaires comme oppressantes.

Près d'un quart (26.75%) étudiants se situent dans une zone intermédiaire, ressentant une pression académique modérée. Ces étudiants peuvent avoir un certain stress lié à leurs études, mais il reste gérable.

Une proportion significative d'étudiants (plus de 40%) ressentent une forte pression académique. Cela pourrait indiquer que de nombreux étudiants sont submergés par leurs charges académiques, ce qui peut avoir des répercussions négatives sur leur bien-être mental et

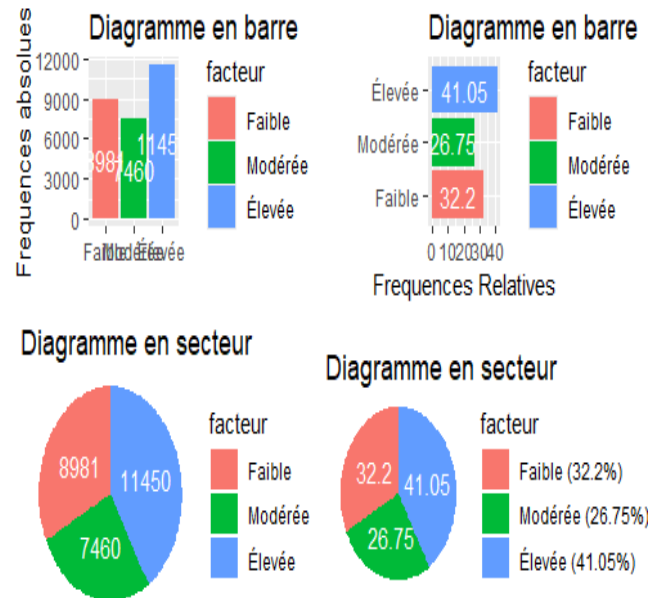


Figure 26 Visualisation de la distribution de la pression académique

5) La variable stress financier

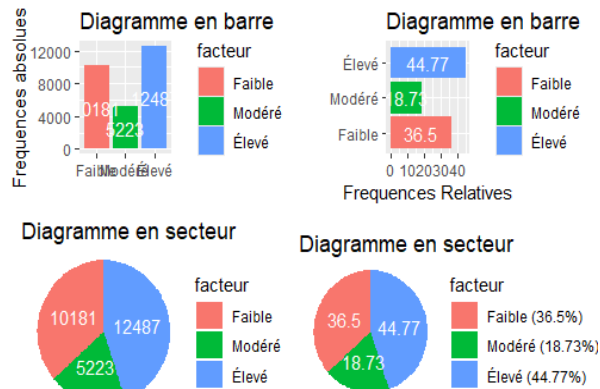


Figure 27 Visualisation de la distribution du stress financier

6) La variable pensée suicidaire

36.71% des étudiants déclarent ne pas avoir de pensées suicidaires. Cela signifie que ces étudiants arrivent à gérer leur stress et leurs difficultés, probablement grâce à des ressources personnelles, un bon soutien social ou un environnement favorable.

Une proportion alarmante (**63.29%**) des étudiants rapporte avoir des pensées suicidaires. Cela révèle une **crise de santé mentale majeure** dans cette population étudiante, qui peut être liée à plusieurs facteurs, notamment la **pression académique élevée**, le **stress financier**, l'**isolement social** ou des **problèmes personnels**.

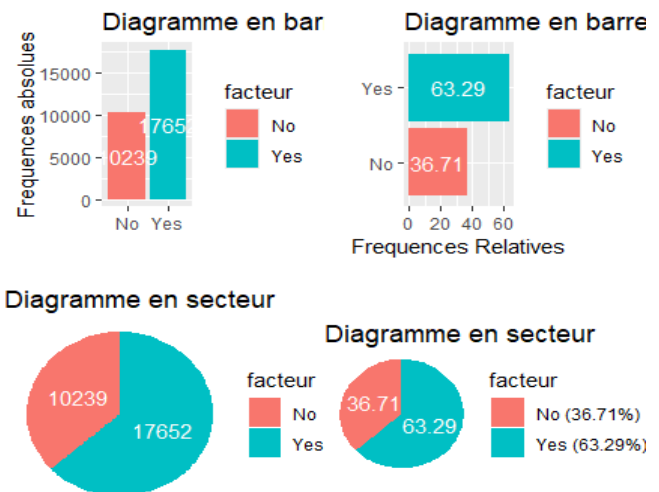


Figure 28 Distribution de la pensée suicidaire

7) La variable habitudes alimentaires

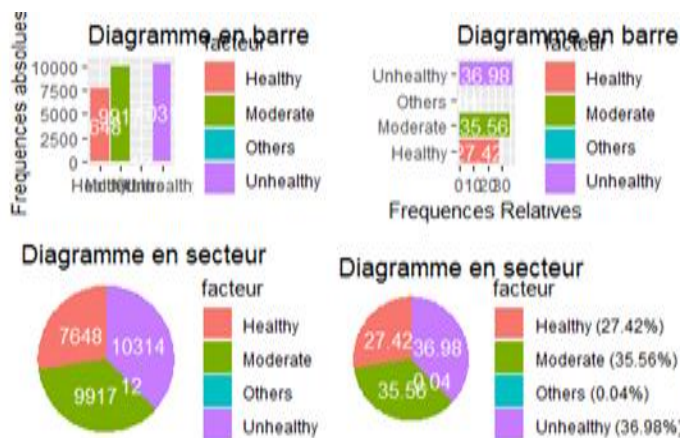


Figure 29 Distribution de l'habitude alimentaire

Un peu plus d'un quart (**27.42%**) des étudiants ont une alimentation saine, ce qui est bénéfique pour leur santé physique et mentale.

35.56% des étudiants alternent entre une alimentation saine et des choix moins équilibrés.

Une proportion **élevée** d'étudiants (presque 37%) suit une alimentation déséquilibrée. Cela peut entraîner des effets négatifs sur leur énergie, leur concentration et même leur santé mentale, augmentant potentiellement le stress et la fatigue.

B) Analyse bivariée

Dans cette partie, nous analysons les relations entre deux variables afin d'identifier d'éventuelles associations ou tendances. En croisant les données, nous mettrons en évidence les liens significatifs et leur impact sur notre étude.

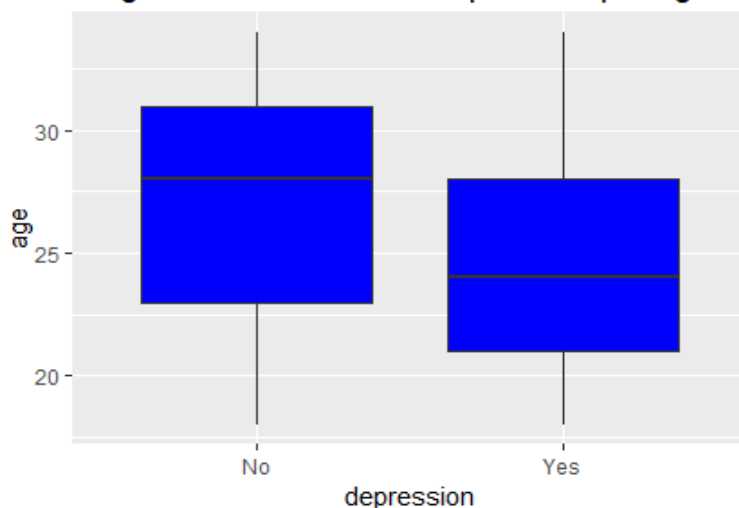
1) Est-ce que dépression chez les étudiants est liée âge ?

Tableau de contingence

Âge	Non	Oui
18	371	1215
19	460	1100
20	657	1578
21	556	1169
22	459	701
23	594	1051
24	748	1509
25	702	1081
26	492	663
27	575	887
28	825	1308
29	845	1103
30	671	474
31	741	686
32	607	654
33	1155	738
34	1101	415

Tableau 6 : Tableau de contingence de la variable dépression et âge.

Diagramme en boîte des depression par âge



Le Rapport de Corrélation entre la dépression et l'âge est **0.05** proche de 0. Cela indique une relation très faible, voire inexistante, entre l'âge et la dépression chez les étudiants. L'âge ne semble donc pas être un facteur déterminant.

Figure 30 Boîte à moustache par niveau de facteur

2) Est-ce que la dépression chez les étudiants dépend du sexe ?

hj

Le V de Cramer est de **0.0019**, ce qui montre une association extrêmement faible entre le sexe et la dépression. Cela suggère que la prévalence de la dépression ne varie pas significativement selon le sexe des étudiants.

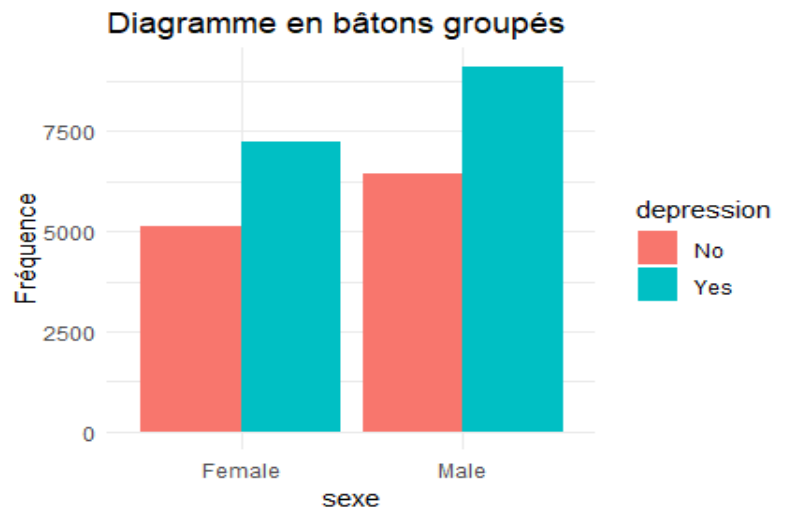
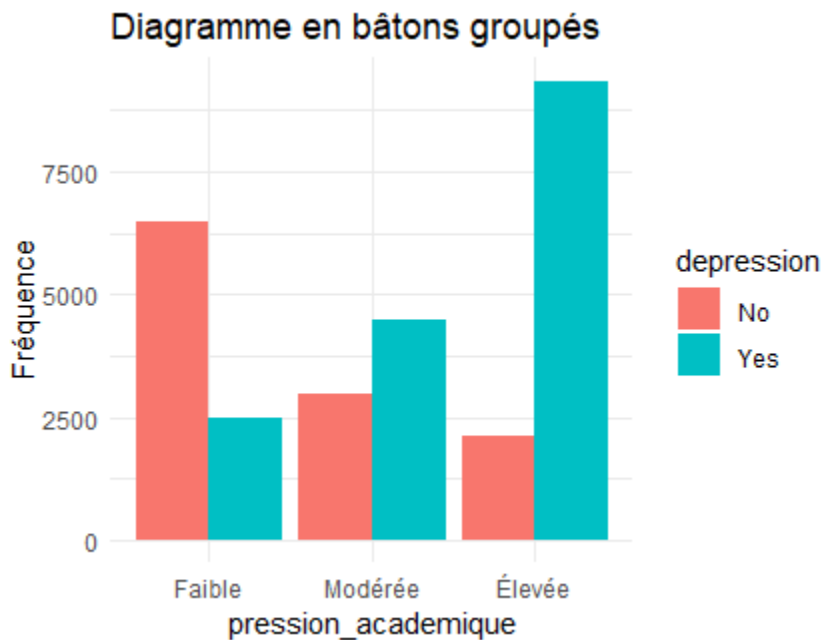


Figure 31 Diagramme en bâtons groupés dépression et sexe

3) Est-ce que la pression dépression chez les étudiants dépend de la pression académique ?



Avec un V de Cramer de **0.4641**, cette relation est la plus forte parmi toutes les variables analysées. La pression académique semble donc jouer un rôle majeur dans l'apparition de la dépression chez les étudiants.

Figure 32 Diagramme en bâtons groupés dépression et pression académique

4) Est-ce que la dépression chez les étudiants est liée au nombre d'heure de travail ou d'étude ?

Le V de Cramer de **0.2148** indique une relation modérée. Cela signifie que la charge de travail ou le temps consacré aux études pourrait avoir un impact significatif sur la dépression des étudiants.

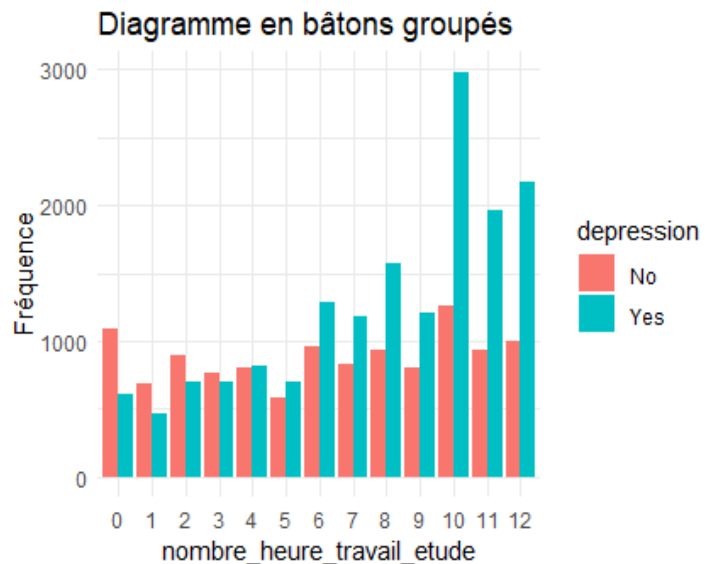
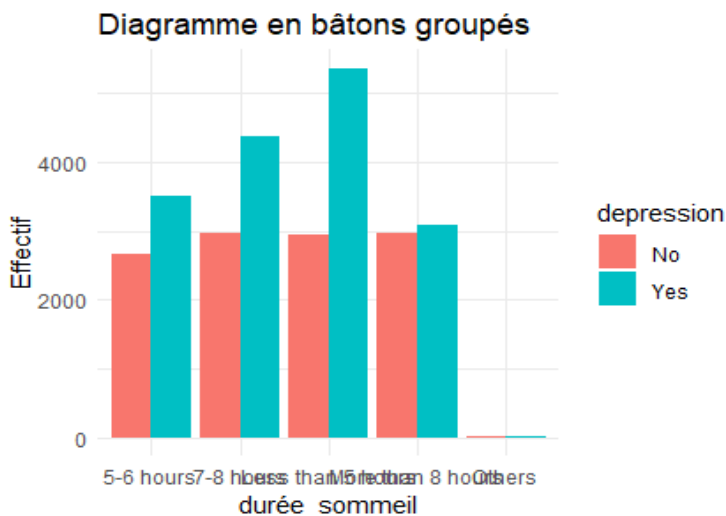


Figure 33 Diagramme en bâtons groupés dépression et nombre_heure_travail_etude

5) Est-ce que la dépression chez les étudiants est causée par la durée de sommeil ?



Un V de Cramer de **0.0995** indique une relation faible, mais non négligeable. Une durée de sommeil insuffisante pourrait être un facteur contributif à la dépression, bien que l'effet ne soit pas très marqué.

Figure 34 Diagramme en bâtons groupés dépression et durée de sommeil

6) Est-ce que la dépression chez les étudiants est liée au stress financier ?

Avec un V de Cramer de **0.0995**, cette relation est également faible, ce qui signifie que le stress financier pourrait jouer un rôle, mais pas de manière prépondérante.

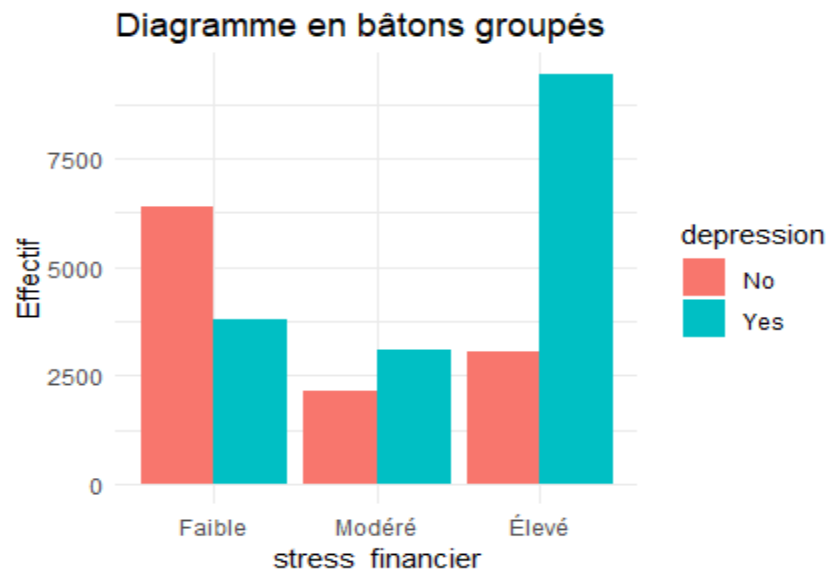


Figure 35 Diagramme en bâtons groupés dépression et stress financier

7) Est-ce que la dépression chez les étudiants est liée aux habitudes alimentaires ?

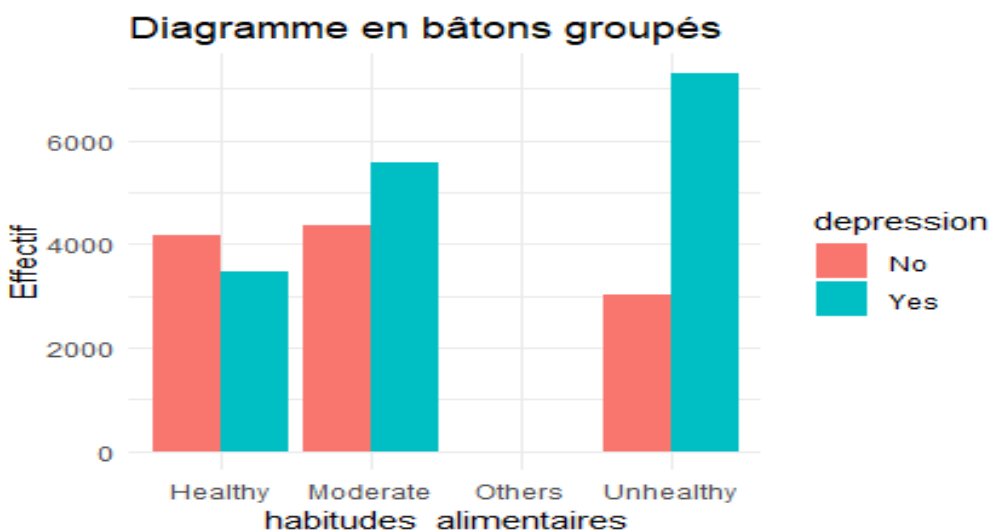


Figure 36 Diagramme en bâtons groupés dépression et habitude alimentaire

Un V de Cramer de **0.2077** indique une relation modérée. Cela suggère que les habitudes alimentaires ont un certain impact sur la dépression, peut-être à travers la qualité de la nutrition et son effet sur l'humeur.

8) Est-ce que la dépression chez les étudiants est liée aux antécédents familiaux de maladie mentale ?

Le V de Cramer de **0.0535** montre une relation faible entre les antécédents familiaux de maladie mentale et la dépression. Bien que des antécédents familiaux puissent être un facteur de risque, ils ne semblent pas jouer un rôle déterminant dans cette population spécifique.

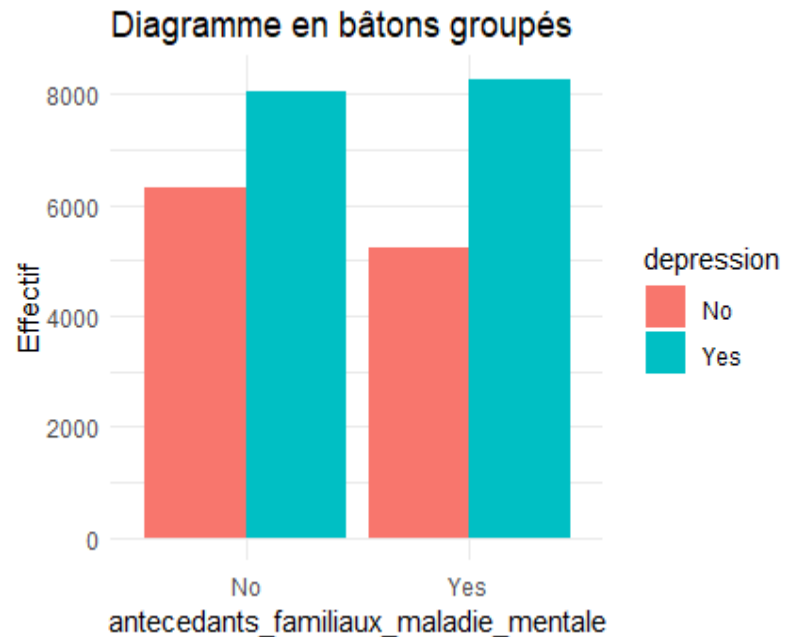
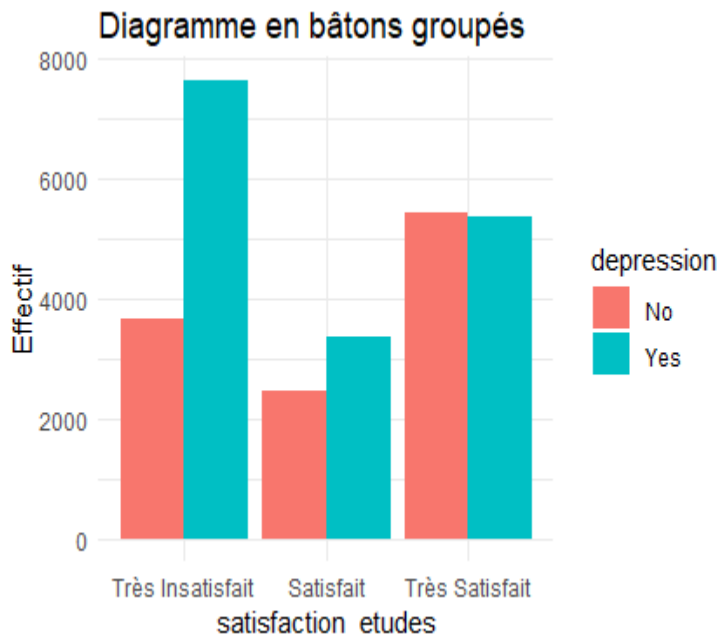


Figure 37 Diagramme en bâtons groupés dépression et antécédents familiaux maladie mentale

9) Est-ce que la dépression chez les étudiants est liée à la satisfaction d'étude ?



Avec un V de Cramer égal à **0.1619** on déduit une relation modérée. Une faible satisfaction dans ses études peut contribuer à la dépression, même si ce n'est pas le facteur le plus fort.

Figure 38 Diagramme en bâtons groupés dépression et satisfaction études

CONCLUSION

Cette étude visait à déterminer les facteurs qui influencent la dépression chez les étudiants, dans un contexte où la santé mentale est de plus en plus au cœur des préoccupations.

Tout d'abord, l'âge, les antécédents familiaux de maladies mentales et le sexe n'ont quasiment pas d'influence sur la dépression. Bien que certaines différences aient été observées entre hommes et femmes, celles-ci ne sont pas assez décisives pour établir une corrélation significative. Cela suggère que la prévention et la prise en charge de la dépression devraient être envisagées de manière inclusive, en tenant compte des besoins individuels plutôt que des distinctions de genre.

En effet, il a été démontré que les jeunes adultes, en fonction du nombre d'heures de travail/étude, la satisfaction d'étude et les habitudes alimentaires présentent des liens modérés, les rendant particulièrement vulnérables à la dépression. Cette relation souligne l'importance d'un accompagnement adapté à quotidien, afin d'offrir un soutien psychologique ciblé aux étudiants les plus à risque.

En revanche la pression académique est le facteur le plus lié à la dépression. Les étudiants confrontés à une charge de travail excessive et à des attentes élevées sont plus susceptibles de développer des symptômes dépressifs. Il est donc essentiel de mettre en place des stratégies de gestion du stress, telles que des programmes de soutien psychologique, des techniques de gestion du temps et une meilleure organisation des évaluations académiques.

Le manque de sommeil et le stress financier ont un impact plus faible, mais non négligeable. Ils jouent également un rôle important dans la vulnérabilité des étudiants à la dépression. Plus la pression financière est forte, plus le risque de souffrir de troubles dépressifs augmente. Face à cette réalité, il est recommandé d'améliorer l'accès aux aides financières, de promouvoir l'éducation budgétaire et d'encourager les initiatives de soutien aux étudiants en difficulté économique.

Comme recommandation nous pouvons citer :

- ✚ Réduire la pression académique par un meilleur accompagnement.
- ✚ Sensibiliser sur l'importance du sommeil et d'une alimentation saine.
- ✚ Soutenir la satisfaction d'étude et la gestion du temps.
- ✚ Proposer un accompagnement financier et psychologique pour les étudiants en difficulté.

Limites de l'analyse

- ✚ Les données peuvent ne pas être représentatives de tous les étudiants (taille ou biais de l'échantillon).
- ✚ Les résultats sont issus d'une analyse bivariée qui ne prouve pas de lien de causalité.
- ✚ D'autres facteurs (comme la personnalité ou l'environnement social) n'ont pas été pris en compte.

Perspectives

- ✚ Réaliser des analyses multivariées (régressions logistiques ou modèles plus complexes) pour mieux cerner l'impact de chaque facteur.
- ✚ Élargir l'échantillon et diversifier les populations étudiées (différentes universités, différents pays).
- ✚ Étudier l'évolution dans le temps (analyses longitudinales) pour mieux comprendre la dynamique de la dépression chez les étudiants.
- ✚ En somme, cette étude met en avant l'importance de la pression académique, de la gestion du temps et de la satisfaction d'étude. Des initiatives ciblées sur ces aspects pourraient aider à réduire la dépression au sein de la population étudiante.

ANNEXES

SOURCE DU CODE << R >>

```
#----- Importation -----  
Student_Depression <- read.csv("C:/Users/DELL/Desktop/First Project Stat/Student_Depression.csv", sep = ";",  
                               stringsAsFactors=TRUE,  
                               row.names = 1)
```

```

Student_Depression$depression <- factor(Student_Depression$depression,
                                         levels = c(0, 1),
                                         labels = c("No", "Yes"))
Student_Depression$pression_liee_au_travail <- factor(Student_Depression$pression_liee_au_travail,
                                                      levels = c(0, 1),
                                                      labels = c("Pas de pression", "Pression"))
Student_Depression$satisfaction_travail <- factor(Student_Depression$satisfaction_travail,
                                                  levels = c(0, 1),
                                                  labels = c("Pas Satisfait", "Satisfait"))
Student_Depression$pression_academique <- cut(Student_Depression$pression_academique,
                                              breaks = c(-Inf, 2, 3, 5), # Définir les intervalles
                                              labels = c("Faible", "Modérée", "Élevée"),
                                              right = TRUE, # Inclure la limite supérieure
                                              ordered_result = TRUE)
Student_Depression$satisfaction_etudes <- cut(Student_Depression$satisfaction_etudes,
                                              breaks = c(-Inf, 2, 3, 5), # Définir les intervalles
                                              labels = c("Très Insatisfait", "Satisfait", "Très Satisfait"),
                                              right = TRUE, # Inclure la limite supérieure
                                              ordered_result = TRUE)
Student_Depression$stress_financier <- cut(Student_Depression$stress_financier,
                                           breaks = c(-Inf, 2, 3, 5), # Définir les intervalles
                                           labels = c("Faible", "Modéré", "Élevé"),
                                           right = TRUE, # Inclure la limite supérieure
                                           ordered_result = TRUE)

print(Student_Depression)
str(Student_Depression)

#Exporter le fichier en local
write.csv(Student_Depression, "C:/Users/DELL/Desktop/INSEDS/Student_Depression.csv",
          row.names = FALSE)

#----- Prétraitement -----

#-----Traitement de doublons
traitement_doublons(Student_Depression)

#-----Traitement des valeurs manquantes
install.packages("visdat")
library(visdat)
vis_dat(Student_Depression)
vis_miss(Student_Depression)
Student_Depression[!complete.cases(Student_Depression),]
nrow(Student_Depression[!complete.cases(Student_Depression),])
nrow(Student_Depression[!complete.cases(Student_Depression),]) / nrow(Student_Depression)
Student_Depression <- na.omit(Student_Depression)
nrow(Student_Depression[!complete.cases(Student_Depression),])

#-----Traitement des valeurs extremes

#----- Visualisation
install.packages('rpart')
library(rpart)
boxplot(Student_Depression$age, col = "red")

#-----Traitement
install.packages('DescTools')
library(DescTools)
Student_Depression$age <- Winsorize(Student_Depression$age)

```

```
boxplot(Student_Depression$age, col = "red")
```

```
#----- Analyse Univariée -----
```

```
# ---- Variable âge
```

```
library(Dagnogo)
```

```
safia.qt.tableau(Student_Depression$age)
```

```
safia.qt.graph(Student_Depression$age)
```

```
install.packages('RVAideMemoire')
```

```
library(RVAideMemoire)
```

```
install.packages("moments")
```

```
library(moments)
```

```
safia.qt.resume(Student_Depression$age)
```

```
# ---- variable nombre_heure_travail_etude
```

```
safia.qt.tableau(Student_Depression$nombre_heure_travail_etude)
```

```
safia.qt.graph(Student_Depression$nombre_heure_travail_etude)
```

```
safia.qt.resume(Student_Depression$nombre_heure_travail_etude)
```

```
#---- variable sexe
```

```
safia.ql.graph(Student_Depression$sexe)
```

```
#---- variable pression_adademique
```

```
safia.ql.graph(Student_Depression$pression_academique)
```

```
#-----stress_financier
```

```
safia.ql.graph(Student_Depression$stress_financier)
```

```
#----- pensée_suicidaire
```

```
safia.ql.graph(Student_Depression$pensees_suicidaire)
```

```
#----- habitudes_alimentaires
```

```
safia.ql.graph(Student_Depression$habitudes_alimentaires)
```

```
#-----Analyse bivariée -----
```

```
# Hypothèse 1: est-ce que la depression chez les étudiants est liée au âge ?
```

```
install.packages('BioStatR')
```

```
library('BioStatR')
```

```
# (Tableau de contingence)
```

```
table(Student_Depression$age, Student_Depression$depression)
```

```
# Graphique: (Boîte par niveau de facteur et horizontalement)
```

```
ggplot(Student_Depression, aes(x = depression, y = age)) +
```

```
  geom_boxplot(fill = "blue") +
```

```
  labs(title = "Diagramme en boîte des depression par âge",
```

```
        x = "depression",
```

```
        y = "age")
```

```
# (Résumé numérique)
```

```
safia.qtpl.liaison(Student_Depression$age, Student_Depression$depression)
```

```
# Hypothèse 2: est-ce que la depression chez les étudiants est liée au sexe ?
```

```
#Tableau
```



```
safia.2ql.tableau(Student_Depression$depression, Student_Depression$sexe)
install.packages('questionr')
safia.2ql.graph(Student_Depression$sexe, Student_Depression$depression)

# (Création d'un DataFrame à partir de la table de contingence)
library(ggplot2)
contingency_table <- table(Student_Depression$depression, Student_Depression$sexe)
contingency_df <- as.data.frame(contingency_table)

# (Création du diagramme en bâtons groupés à l'aide de...)
ggplot(contingency_df, aes(x = Var2, y = Freq,
                           fill = Var1)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  xlab("sexe") +
  ylab("Fréquence") +
  ggtitle("Diagramme en bâtons groupés") +
  guides(fill = guide_legend(title = "depression"))

# (Résumé numérique)
safia.2ql.liaison(Student_Depression$sexe, Student_Depression$depression)

# Hypothèse 3: est-ce que la depression chez les étudiants est liée à la pression academique ?
safia.2ql.tableau(Student_Depression$depression, Student_Depression$pression_academique)

# Graphique
# (Création d'un DataFrame à partir de la table de contingence)
library(ggplot2)
contingency_table <- table(Student_Depression$depression, Student_Depression$pression_academique)
contingency_df <- as.data.frame(contingency_table)

# (Création du diagramme en bâtons groupés à l'aide de...)
ggplot(contingency_df, aes(x = Var2, y = Freq,
                           fill = Var1)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  xlab("pression_academique") +
  ylab("Fréquence") +
  ggtitle("Diagramme en bâtons groupés") +
  guides(fill = guide_legend(title = "depression"))

# Résumé
safia.2ql.liaison(Student_Depression$depression, Student_Depression$pression_academique)
library(questionr)
cramer.v(table(Student_Depression$depression, Student_Depression$pression_academique))

# Hypothèse 4: Est-ce que la dépression chez les étudiants est liée au nombre d'heure de travail ou d'étude ?
# Graphique
library(ggplot2)
contingency_table <- table(Student_Depression$depression, Student_Depression$nombre_heure_travail_etude)
contingency_df <- as.data.frame(contingency_table)

# (Création du diagramme en bâtons groupés à l'aide de...)
ggplot(contingency_df, aes(x = Var2, y = Freq,
                           fill = Var1)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  xlab("nombre_heure_travail_etude") +
  ylab("Fréquence") +
```

```
ggtitle("Diagramme en bâtons groupés") +
guides(fill = guide_legend(title = "depression"))

# Résumé
cramer.v(table(Student_Depression$depression, Student_Depression$nombre_heure_travail_etude))

# Hypothèse 5: Est-ce que la dépression chez les étudiants est causée par la durée de sommeil ?
# Graphique
library(ggplot2)
contingency_table <- table(Student_Depression$depression, Student_Depression$duree_sommeil)
contingency_df <- as.data.frame(contingency_table)

# (Création du diagramme en bâtons groupés à l'aide de...)
ggplot(contingency_df, aes(x = Var2, y = Freq,
                           fill = Var1)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  xlab("durée_sommeil") +
  ylab("Effectif") +
  ggtitle("Diagramme en bâtons groupés") +
  guides(fill = guide_legend(title = "depression"))

# Résumé
cramer.v(table(Student_Depression$depression, Student_Depression$duree_sommeil))

# Hypothèse 6: Est-ce que la dépression chez les étudiants est liée au stress financier ?
# Graphique
contingency_table <- table(Student_Depression$depression, Student_Depression$stress_financier)
contingency_df <- as.data.frame(contingency_table)

# (Création du diagramme en bâtons groupés à l'aide de...)
ggplot(contingency_df, aes(x = Var2, y = Freq,
                           fill = Var1)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  xlab("stress_financier") +
  ylab("Effectif") +
  ggtitle("Diagramme en bâtons groupés") +
  guides(fill = guide_legend(title = "depression"))

# Résumé
cramer.v(table(Student_Depression$depression, Student_Depression$duree_sommeil))

# Hypothèse 7: Est-ce que la dépression chez les étudiants est liée aux habitudes alimentaires ?
# Graphique
contingency_table <- table(Student_Depression$depression, Student_Depression$habitudes_alimentaires)
contingency_df <- as.data.frame(contingency_table)

# (Création du diagramme en bâtons groupés à l'aide de...)
ggplot(contingency_df, aes(x = Var2, y = Freq,
                           fill = Var1)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  xlab("habitudes_alimentaires") +
  ylab("Effectif") +
  ggtitle("Diagramme en bâtons groupés") +
  guides(fill = guide_legend(title = "depression"))
```

Résumé

```
cramer.v(table(Student_Depression$depression, Student_Depression$habitudes_alimentaires))
```

Hypothèse 8: Est-ce que la dépression chez les étudiants est liée aux antécédants familiaux de maladie mentale ?

```
contingency_table <- table(Student_Depression$depression, Student_Depression$antecedants_familiaux_maladie_mentale)
contingency_df <- as.data.frame(contingency_table)
```

(Création du diagramme en bâtons groupés à l'aide de...)

```
ggplot(contingency_df, aes(x = Var2, y = Freq,
                           fill = Var1)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  xlab("antecedants_familiaux_maladie_mentale") +
  ylab("Effectif") +
  ggtitle("Diagramme en bâtons groupés") +
  guides(fill = guide_legend(title = "depression"))
```

Résumé

```
cramer.v(table(Student_Depression$depression, Student_Depression$antecedants_familiaux_maladie_mentale))
```

Hypothèse 9: Est-ce que la dépression chez les étudiants est liée à la satisfaction d'étude ?

```
contingency_table <- table(Student_Depression$depression, Student_Depression$satisfaction_etudes)
contingency_df <- as.data.frame(contingency_table)
```

(Création du diagramme en bâtons groupés à l'aide de...)

```
ggplot(contingency_df, aes(x = Var2, y = Freq,
                           fill = Var1)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  xlab("satisfaction_etudes") +
  ylab("Effectif") +
  ggtitle("Diagramme en bâtons groupés") +
  guides(fill = guide_legend(title = "depression"))
```

Résumé

```
cramer.v(table(Student_Depression$depression, Student_Depression$satisfaction_etudes))
```