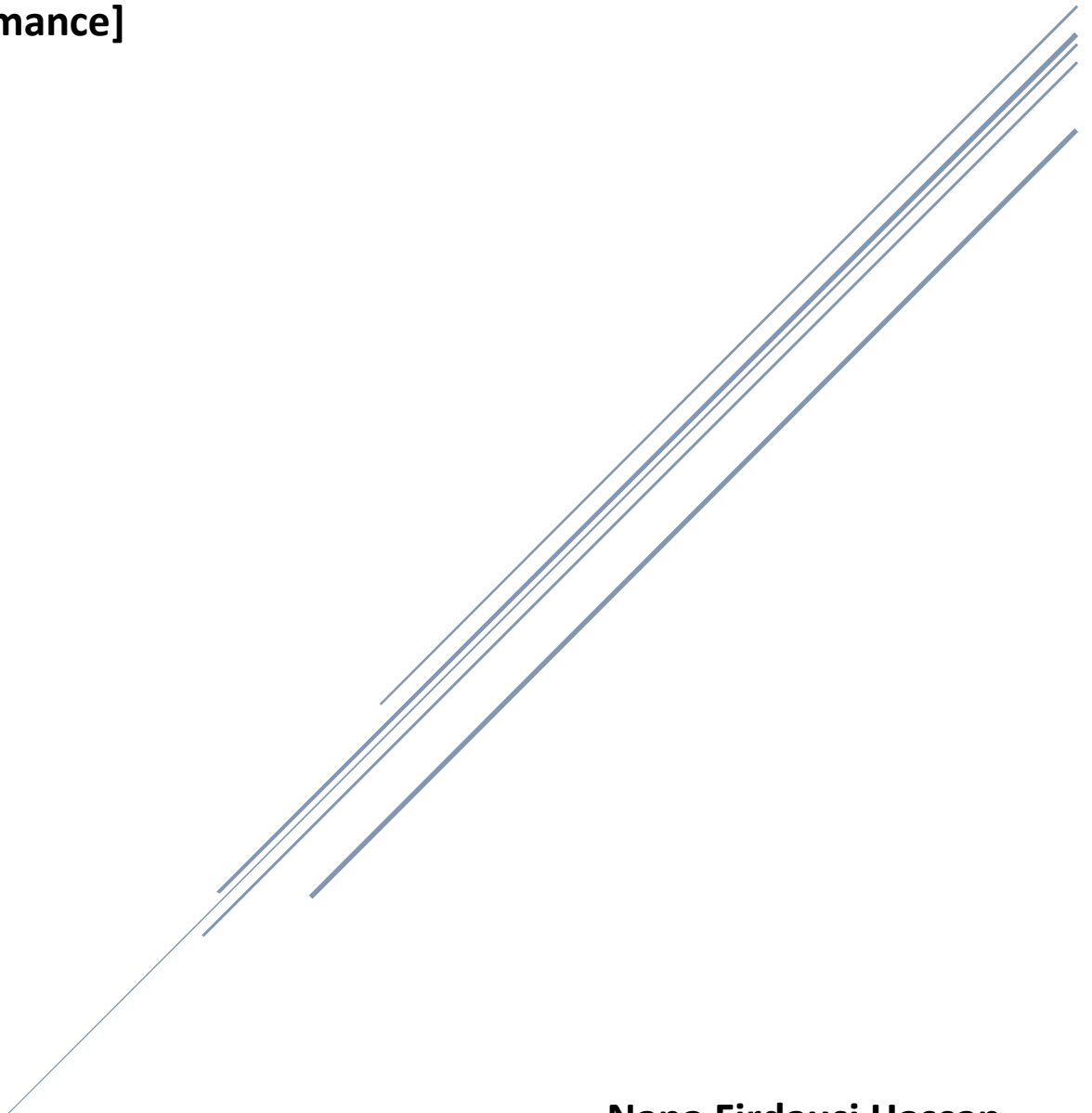


MUltifactorial influences on university students' CGPA.

[Examining the Impact of Age, Course, Gender, Sleep Quality, Diet, Relationships, Substance Use, Counseling, Family History, and Chronic Illness on Academic Performance]



Nana Firdausi Hassan

Q1. Research question.

How do various factors, including age, course of study, gender, sleep quality, diet quality, relationship status, substance use, counseling service use, family history, and chronic illness, contribute to variations in students' CGPA in a university setting. (I aim to investigate the multifaceted influences on students' academic performance, as measured by CGPA).

Q2. Information about where each column of data came from.

Students Mental Health Assessments.

https://www.kaggle.com/datasets/sonia22222/students-mental-health-assessments#:~:text=calendar_view_week-,students_mental_health_survey,-.csv
[Final project.xlsx](#)

Q3. The descriptive statistics table includes all independent and dependent variables.

Variables	Mean	Median	Standard Deviation	Kurtosis	Skewness	Range	Minimum	Maximum	Count
CGPA	3.482	3.5	0.324	25.510	-2.587	4	0	4	1499
Age	1.243	1	0.429	-0.559	1.201	1	1	2	1499
Course	3.127	3	1.690	-1.021	-0.490	5	0	5	1499
Gender	0.476	0	0.500	-1.994	0.095	1	0	1	1499
SleepQlty	1.432	2	0.663	-0.531	-0.749	2	0	2	1499
DietQlty	1.000	1	0.607	-0.281	0.000	2	0	2	1499
RltshpStatus	0.680	0	0.782	-1.096	0.628	2	0	2	1499
SubsUse	0.612	0	6.263	241.936	15.550	99	0	99	1499
CnslngSerUse	0.511	0	0.675	-0.283	0.964	2	0	2	1499
FmlyHistory	0.308	0	0.462	-1.305	0.835	1	0	1	1499
ChrnclIllness	0.048	0	0.214	15.927	4.232	1	0	1	1499

Q4. A correlation coefficients table that includes all independent and dependent variables.

<i>Variables</i>	<i>CGPA</i>	<i>Age</i>	<i>Course</i>	<i>Gender</i>	<i>SlpQty</i>	<i>DietQty</i>	<i>RltpStat</i>	<i>SubsUse</i>	<i>CnselngUse</i>	<i>FmlyHstry</i>	<i>ChrnIll</i>
CGPA	1										
Age	0.00	1									
Course	-0.02	-0.03	1								
Gender	0.04	-0.05	0.00	1							
SleepQty	0.00	0.03	-0.01	0.01	1						
DietQty	-0.01	0.01	0.05	-0.04	-0.03	1					
RltshpStatus	0.00	-0.02	0.00	-0.02	-0.19	-0.14	1				
SubsUse	0.02	0.04	0.00	-0.02	0.01	0.00	0.03	1			
CnselngSerUse	0.00	-0.01	0.00	0.00	-0.07	0.03	0.00	-0.02	1		
FmlyHistory	0.01	-0.02	0.01	-0.01	0.02	0.01	-0.03	-0.02	0.01	1	
Chrncllness	-0.02	-0.03	0.01	0.00	-0.03	-0.01	-0.01	0.04	-0.01	0.01	1

Q5. Explanation of any multicollinearity concerns and how they were addressed in the multiple regression.

Based on the correlation matrix provided, there doesn't seem to be a significant multicollinearity concern among the independent variables. Therefore, no further steps to address multicollinearity were necessary.

Q6. The three multiple regression output tables (does include the residual table, just the regression statistics, ANOVA, and regression table with coefficients and p-values).

SUMMARY
OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.05391
R Square	0.00291
Adjusted R Square	-0.00379
Standard Error	0.32415
Observations	1499

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	10	0.45577	0.0455	0.4337	0.93056
Residual	1488	156.35356	0.1050		
Total	1498	156.80934			

	Coefficien ts	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.483	0.042	82.568	0	3.401	3.566	3.401	3.566
Age	0.001	0.020	0.069	0.945	-0.037	0.040	-0.037	0.040
Course	-0.004	0.005	-0.848	0.396	-0.014	0.006	-0.014	0.006
Gender	0.026	0.017	1.522	0.128	-0.007	0.059	-0.007	0.059
SleepQlty	0.000	0.013	-0.014	0.989	-0.026	0.025	-0.026	0.025
DietQlty	-0.005	0.014	-0.376	0.707	-0.033	0.022	-0.033	0.022
RltshpStatus	0.001	0.011	0.079	0.937	-0.021	0.023	-0.021	0.023
SubsUse	0.001	0.001	0.702	0.483	-0.002	0.004	-0.002	0.004
CnseIngSerUse	0.002	0.012	0.198	0.843	-0.022	0.027	-0.022	0.027
FmlyHistory	0.008	0.018	0.437	0.662	-0.028	0.044	-0.028	0.044
ChrncIllness	-0.026	0.039	-0.659	0.510	-0.103	0.051	-0.103	0.051

Q7. For each independent variable, I provide (a) an alternative hypothesis (Note: Some of the hypotheses are non-directional, and some are directional), (b) a rationale for the alternative hypothesis (e.g., why is it directional/non-directional?), (c) a null hypothesis, (d) the exact p-value based on the hypothesis (this is a specific number), (e) a conclusion about the null hypothesis, and (f) an interpretation of the coefficient (including units of measurement for the independent and dependent variable).

Age

- a. Alternative Hypothesis H1 (Non-Directional): There is a significant difference in CGPA between older and younger students.
- b. Rationale: I assume there is a significant difference in CGPA and age without predicting whether older students will have a higher or lower CGPA than younger students.
- c. Null Hypothesis (H0): There is no significant difference in CGPA between older and younger students.
- d. 0.945
- e. Fail to reject the null hypothesis
- f. Interpretation: For every one-unit increase in the age variable, the CGPA increases by 0.001 units. The p-value 0.945 suggests that this coefficient is not statistically significant. There is no statistically significant difference in the CGPA between age groups.

Course

- a. Alternative Hypothesis H1 (Directional): Students in Business courses have a significantly higher CGPA compared to students in other courses.
- b. Rationale: I predicted and expect that Business students will have a higher CGPA compared to other students.
- c. Null Hypothesis (H0): Students in Business courses do not have a significantly higher CGPA compared to students in other courses.
- d. 0.396
- e. Fail to reject the null hypothesis.
- f. Interpretation: For every one-unit increase in the course variable, the CGPA decreases by 0.004 units. The p-value 0.396 suggests that this coefficient is not statistically significant. There is no statistically significant difference in the CGPA and course.

Gender

- a. Alternative Hypothesis H1 (Directional): Male students have a significantly higher CGPA compared to female students.
- b. Rationale: I predicted and expect that Male students will have a higher CGPA compared to female students.
- c. Null Hypothesis (H0): Male students do not have a significantly higher CGPA compared to female students.
- d. 0.128
- e. Fail to reject the null hypothesis.
- f. Interpretation: For every one-unit increase in the gender variable, the CGPA increases by 0.026 units. The p-value 0.128 suggests that this coefficient is not statistically significant. There is no statistically significant difference in the CGPA and gender.

Sleep Quality

- a. Alternative Hypothesis (Non-directional): There is a statistically significant difference between sleep quality of student and CGPA
- b. Rationale: I assume there is a significant difference in CGPA and sleep quality without predicting whether students with good sleep quality have higher CGPA than those without good sleep quality.
- c. Null Hypothesis (H0): There is no statistically significant difference between sleep quality of student and CGPA
- d. 0.989
- e. Fail to reject the null hypothesis
- f. Interpretation: For every one-unit increase in the sleep quality variable, the CGPA increases by 0.000 units. The p-value 0.989 suggests that this coefficient is not statistically significant. There is no statistically significant difference between CGPA and sleep quality.

Diet Quality

- a. Alternative Hypothesis (Non-directional): There is a statistically significant difference between diet quality of student and CGPA
- b. Rationale: I assume there is a significant difference in CGPA and diet quality without predicting whether students with poor diet quality have lower CGPA than those with good diet quality.
- c. Null Hypothesis (H0): There is no statistically significant difference between diet quality of student and CGPA
- d. 0.707
- e. Fail to reject the null hypothesis
- f. Interpretation: For every one-unit increase in the diet quality variable, the CGPA decreases by 0.005 units. The p-value 0.707 suggests that this coefficient is not statistically significant. There is no statistically significant difference between CGPA and diet quality.

Relationship Status

- a. Alternative Hypothesis (Non-directional): There is a significant difference between relationship status and CGPA of students.
- b. Rationale: I assume there is a significant difference between CGPA and relationship status without predicting whether students in a relationship have higher/lower CGPA than those that are not.
- c. Null Hypothesis (H0): There is no significant difference between relationship status and CGPA of students.
- d. 0.937
- e. Fail to reject the null hypothesis
- f. Interpretation: For every one-unit increase in the relationship status variable, the CGPA increases by 0.001 units. The p-value 0.937 suggests that this coefficient is not statistically significant. There is no statistically significant difference between CGPA and relationship status.

Substance Use

- a. Alternative Hypothesis (Directional): Students who engage in substance use have a significantly lower CGPA.
- b. Rationale: I predicted and expected that students who use substances (i.e. can negatively impact cognitive function, attention, and academic engagement) have lower CGPA compared to students who don't.
- c. Null Hypothesis (H0): Students who engage in substance use do not have a significantly lower CGPA.
- d. 0.483
- e. Fail to reject the null hypothesis

f. Interpretation: For every one-unit increase in the substance use variable, the CGPA increases by 0.001 units. The p-value 0.483 suggests that this coefficient is not statistically significant. There is no statistically significant difference between CGPA and substance use.

Counseling Service Use

a. Alternative Hypothesis H1 (Directional): Students who utilize counseling services have a significantly higher CGPA.

b. Rationale: I predicted and expected that students who utilize counseling services have higher CGPA compared to students who don't.

c. Null Hypothesis (H0): Students who utilize counseling services have a significantly lower CGPA

d. 0.843

e. Fail to reject the null hypothesis

f. Interpretation: For every one-unit increase in the counseling service use variable, the CGPA increases by 0.002 units. The p-value 0.843 suggests that this coefficient is not statistically significant. There is no statistically significant difference between CGPA and counseling service use.

Family History

a. Alternative Hypothesis (Non-directional): There is a statistically significant relationship between family history and students CGPA.

b. Rationale: I assume there is a significant relationship between CGPA and family history without predicting whether students with positive family influences, such as a history of academic success, may contribute to higher CGPA.

c. Null Hypothesis (H0): There is no statistically significant relationship between family history and students' CGPA.

d. 0.662

e. Fail to reject the null hypothesis

f. Interpretation: For every one-unit increase in the family history variable, the CGPA increases by 0.008 units. The p-value 0.662 suggests that this coefficient is not statistically significant. There is no statistically significant difference between CGPA and family history.

Chronic Illness

a. Alternative Hypothesis H1 (Directional): Students with a chronic illness have a significantly lower CGPA.

b. Rationale: I predicted and expect that students with chronic illness will have a lower CGPA compared to other students who do not have.

c. Null Hypothesis (H0): Students with a chronic illness have a significantly higher CGPA.

d. 0.510

e. Fail to reject the null hypothesis

f. Interpretation: For every one-unit increase in the chronic illness variable, the CGPA decreases by 0.026 units. The p-value 0.510 suggests that this coefficient is not statistically significant. There is no statistically significant difference between CGPA and chronic illness.

Q8. For the overall regression model, (a) explanation on the R², (b) the regression equation, and (c) selection of an actual observation (i.e., row of data) to calculate what the regression equation would estimate

the dependent variable to be based on the independent variables in that row of data. Was the estimate close to the actual value for the dependent variable in that row of data?

a. R square: The model explains only 0.29% of the variation in the CGPA

b. Regression equation:

$$= 3.483 + 0.001 \times \text{Age} - 0.004 \times \text{Course} + 0.026 \times \text{Gender} - 0.000 \times \text{SleepQuality} - 0.005 \times \text{DietQuality} + 0.001 \times \text{RelationshipStatus} + 0.001 \times \text{SubstanceUse} + 0.002 \times \text{CounselingServiceUse} + 0.008 \times \text{FamilyHistory} - 0.026 \times \text{ChronicIllness}$$

c. Estimate of regression equation in row 319

CGPA	Age	Course	Gender	Sleep_Quality	Diet_Quality	Relationship_Status	Substance_Use	Counseling_Service_Use	Family_History	Chronic_Illness
2.84	1	5	0	2	2	0	0	0	0	0

$$\begin{aligned}
 &= 3.483 + 0.001 \times 1 - 0.004 \times 5 + 0.026 \times 0 + 0.000 \times 2 - 0.005 \times 2 + 0.001 \times 0 + 0.001 \times 0 + 0.002 \times 0 + 0.008 \times 0 - 0.026 \times 0 \\
 &= 3.483 + 0.001 - 0.02 + 0 + 0 - 0.01 + 0 + 0 + 0 + 0 - 0 \\
 &= 3.454
 \end{aligned}$$

Actual CGPA in the row is 2.84

The estimate value of Severity of CGPA 3.454 is quite close to the actual value of 2.84 for this specific Observation with a difference of 0.614.

Q9. Without using statistics jargon, what can you conclude from the multiple regression? Relate the results back to your answer to #1 and discuss any other interesting findings.

According to the result on this multiple regression, it is revealed that several factors influence students' CGPA in university settings, although none are significant, demonstrating a complicated interaction between personal and environmental factors. Additional research is required to uncover particular causes. The multiple regression analysis, based on the provided data, does not support a strong relationship between the chosen factors and students' CGPA. The model lacks predictive power, as evidenced by the low R² and non-significant coefficients. It is important to consider that other unmeasured factors may influence academic performance, and further investigation or refinement of the model may be necessary to better understand the multifaceted influences on students' CGPA.