

STATISTICAL INFERENCE FOR DECISION MAKING

ANALYSIS OF DIFFERENT FACTORS THAT IMPACT CGPA

Submitted to :

Prof.(Dr.) Sridhar Vaithianathan

GROUP 1



Srushti Jagwani
B004
80662400119



Amisha Aggarwal
B014
80662400182



Ishika Rupareliya
B024
80662400132



Ansh Kumar
B034
80662400152



Prajay Singh
B044
80662400163



Ashwin Kumar Gondi
B054
80662400047



Vaibhav
B064
80662400091

TABLE OF CONTENT

- Background
- Data Definition
- Two sample Z-Test
- One sample T-Test
- Two sample T-Test(unequal variance)
 - Two sample T-Test(paired means)
 - ANOVA
 - Chi-square
- Multiple linear regression
- Multiple linear regression output analysis

BACKGROUND



Students' academic performance is influenced by a multitude of factors, one of which is the time and effort spent traveling from home to college. The time taken to commute, the mode of transportation, and the distance between home and college may affect students' study habits, participation in extracurricular activities, and overall academic performance. This project aims to analyze these factors to determine if there is a significant impact on students' academic outcomes.

Purpose: The purpose of this project is to statistically assess the relationship between travel-related factors (such as distance from college, travel time, and mode of commute) and students' academic performance. The study will utilize various statistical techniques to analyze the data and draw meaningful conclusions .

DATA DEFINITION

Source: The dataset is a simulated collection of 120 entries representing students' travel and lifestyle characteristics and their academic performance, created specifically for this analysis.

Data Definition:

- **Distance (km):** The distance from the student's home to the college.
- **Time (minutes):** The time taken to travel from home to college.
- **Travel Commute:** The mode of transportation used by the student. 3 main mode of transport used are Auto, Local Train and Uber/Ola.
- **CGPA:** Cumulative Grade Point Average of the student, scaled to a maximum of 4.
- **Self-study time (hours/day):** The average daily time spent by students in self-study.
- **Clubs Joined:** The number of extracurricular clubs joined by the student.(maximum 3)
- **Events Participated:** The number of events the student has participated in after the classes.
- **Living Situation:** The student's living arrangement (with family, hostel, rented flat, PG).
- **Study Time Before Joining Clubs:** Daily hours spent in self-study before joining clubs.
- **Study Time After Joining Clubs:** Daily hours spent in self-study after joining clubs.
- **Sleep Hours (hrs/day):** The average hours of sleep per day.
- **Exhaustion after Traveling:** The level of exhaustion experienced after traveling (very, ok, not at all).
- **Social Media Usage (hrs/day):** The average daily time spent on social media in hrs.

TWO SAMPLE Z-TEST

Objective of the Test: To compare the time taken by the students by local train and cab.

Test Applied: Two Sample Independent Z-Test

Hypothesis Tested:

- **Null Hypothesis (H_0):** There is no significant difference in the time taken by student travelling through cab and local train.
- **Alternative Hypothesis (H_1):** There is significant difference in the time taken by student travelling through cab and local train.

Calculations:

- Perform the two sample z-test using Excel's Data Analysis Toolpak.
- Obtain the z-value and the p-value.

Decision Taken:

- P value = 0.449494983347066
- p critical= 0.05
- p value > 0.05 , Fail to reject H_0

Managerial Inference:

H_0 is not rejected , so there is no significant difference in the time taken by student travelling through cab and local train.

z-Test: Two Sample for Means		
	<i>Time (minutes) Local Train</i>	<i>Time (minutes) Cab</i>
Mean	30.33333333	31.29545455
Known Variance	62.45877	17.32928
Observations	51	44
Hypothesized Mean Difference	0	
z	-0.756257234	
P(Z<=z) one-tail	0.224747492	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.449494983	
z Critical two-tail	1.959963985	

ONE SAMPLE T-TEST

Objective of the Test: To determine if the average CGPA of the student significantly differs from a average mean CGPA of 3.

Test Applied: One Sample T-Test on CGPA of the students.

Hypothesis Tested:

- **Null Hypothesis (H_0):** The average CGPA of students is equal to the hypothesized mean CGPA of 3.0.
- **Alternative Hypothesis (H_1):** The average CGPA of students is not equal to the hypothesized mean CGPA of 3.0.

Calculations:

- Perform the one sample t-test using Excel's Data Analysis Toolpak.
- Obtain the t-value and the p-value.

Decision Taken:

- P value =0.05361009
- p critical= 0.05
- p value > 0.05 , Fail to reject H_0

Managerial Inference:

H_0 is not rejected , so there is no significant difference between the average CGPA of the students and the hypothesized mean CGPA of 3

CGPA	
Mean	3.111583333
Standard Error	0.027074437
Median	3.09
Mode	2.78
Standard Deviation	0.2965856
Sample Variance	0.087963018
Kurtosis	-0.580198495
Skewness	0.031063786
Range	1.32
Minimum	2.44
Maximum	3.76
Sum	373.39
Count	120
Confidence Level(95.0%)	0.05361009

TWO SAMPLE T-TEST

Objective of the Test: To determine if there's a significant difference in CGPA between student who live less than 4km away from the college and those who live more than 4Km away from the college.

Test Applied: Two Sample Independent T-Test of unequal variance

Hypothesis Tested:

- **Null Hypothesis (H_0):** There is no significant difference in CGPA of student who live less than 4km away from the college and those who live more than 4Km away from the college.
- **Alternative Hypothesis (H_1):** There is a significant difference in CGPA between the two groups.

Calculations:

- Perform the two sample t-test using Excel's Data Analysis Toolpak.
- Obtain the t-value and the p-value.

Decision Taken:

- Tstat = 14.0487207372708
- P value = 1.28068395152727E-24
- p critical= 0.05
- p value < 0.05 , Reject H_0
- t stat > t critical , Reject H_0

Managerial Inference:

H_0 is rejected , so there is significant difference in CGPA of student who live less than 4km away from the college and those who live more than 4Km away from the college.

t-Test: Two-Sample Assuming Unequal Variances			
	<4	>4	
Mean	3.23516854	2.75677419	
Variance	0.05322071	0.01740925	
Observations	89	31	
Hypothesized	0		
df	92		
t Stat	14.0487207		
P(T<=t) one-ta	6.4034E-25		
t Critical one-	1.6615854		
P(T<=t) two-ta	1.2807E-24		
t Critical two-	1.98608632		

TWO SAMPLE T-TEST

Objective of the Test: To analyze whether joining clubs has significantly affected students' CGPA.

Test Applied: Two Sample T-Test (Paired two sample for means).

Hypothesis Tested:

- **Null Hypothesis (H_0):** There is no difference in CGPA of student before and after joining clubs.
- **Alternative Hypothesis (H_1):** There is a significant difference in CGPA of student before and after joining clubs.

Calculations:

- Perform the two sample t-test using Excel's Data Analysis Toolpak.
- Obtain the t-value and the p-value.

Decision Taken:

- t stat = 10.9152073293313
- P value = 1.21186516659793E-19
- p critical= 0.05
- p value < 0.05 , Reject H_0
- t stat > t critical , Reject H_0

Managerial Inference:

H_0 is rejected , so there is significant difference in CGPA of student before and after joining clubs.

t-Test: Paired Two Sample for Means			
	Study Time Before Join	ime After Joining Clubs	
Mean	2.175	1.49166667	
Variance	0.24642857	0.25203081	
Observations	120	120	
Pearson Corre	0.05648011		
Hypothesized	0		
df	119		
t Stat	10.9152073		
P(T<=t) one-ta	6.0593E-20		
t Critical one-	1.65775928		
P(T<=t) two-ta	1.2119E-19		
t Critical two-	1.98009988		

ANOVA

Objective of the Test: To determine if there is a significant difference in CGPA among students using different modes of travel (Auto, Local Train and Uber/Ola).

Test Applied: One way ANOVA.

Hypothesis Tested:

- **Null Hypothesis (H_0):** There is no significant difference in CGPA among students using different modes of travel (Auto, Local Train and Uber/Ola).
- **Alternative Hypothesis (H_1):** There is a significant difference in CGPA among students using different modes of travel (Auto, Local Train and Uber/Ola).

Calculations:

- Perform the ANOVA test using Excel's Data Analysis Toolpak.
- Obtain the f-value and the p-value.

Decision Taken:

- $F = 59.0345084191523$
- $P \text{ value} = 1.87906820273081E-18$
- $p \text{ critical} = 0.05$
- $p \text{ value} < 0.05$, Reject H_0
- $F > F \text{ critical}$, Reject H_0

Managerial Inference:

H_0 is rejected, so there is a significant difference in CGPA among students using different modes of travel (Auto, Local Train and Uber/Ola).

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Auto	68	223.42	3.28558824	0.05271159		
Local Train	41	116.15	2.83292683	0.02480622		
Uber/ola	11	33.82	3.07454545	0.06860727		
ANOVA						
Source of Variati	SS	df	MS	F	P-value	F crit
Between Grou	5.25760119	2	2.62880059	59.0345084	1.88E-18	3.0737629
Within Group	5.20999798	117	0.0445299			
Total	10.4675992	119				

ANOVA

Objective of the Test: To determine if there is a significant difference in CGPA among students living in different environments (Hostels, PG, Rented Flats and with family).

Test Applied: One way ANOVA.

Hypothesis Tested:

- **Null Hypothesis (H_0):** The CGPA is the same across all living situations.
- **Alternative Hypothesis (H_1):** At least one group has a significantly different mean CGPA.

Calculations:

- Perform the ANOVA test using Excel's Data Analysis Toolpak.
- Obtain the f-value and the p-value.

Decision Taken:

- $F = 24.9918578246935$
- $P \text{ value} = 1.51833994817332E-12$
- $p \text{ critical} = 0.05$
- $p \text{ value} < 0.05$, Reject H_0
- $F > F \text{ critical}$, Reject H_0

Managerial Inference:

H_0 is rejected, so there is a significant difference in CGPA among students living in different environment. A significant result suggest that living conditions affect academic performance

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Hostel	38	126.36	3.3252632	0.0726526		
PG	27	86.13	3.19	0.05		
Rented Flats	45	133.22	2.9604444	0.0478907		
With Family	10	27.68	2.768	0.0291956		
ANOVA						
Source of Variati	SS	df	MS	F	P-value	F crit
Between Grou	4.1095007	3	1.3698336	24.991858	1.52E-12	2.6828094
Within Group	6.3580985	116	0.0548112			
Total	10.467599	119				

CHI-SQUARE

Objective of the Test: To examine if there is an association between the mode of commute and the level of exhaustion after traveling

Test Applied: Chi-Square Test of Independence.

Hypothesis Tested:

- Null Hypothesis (H_0): Exhaustion level is independent of the travel mode. (There is no association between travel mode and exhaustion level.)
- Alternative Hypothesis (H_1): Exhaustion level is dependent of the travel mode. (There is a significant association between travel mode and exhaustion level.)

Calculations:

- Calculate the Chi-Square statistic using the formula: $\chi^2 = \sum (O_i - E_i)^2 / E_i$
- where O_i is the observed frequency and E_i is the expected frequency.
- Obtain the Chi-square value and the p-value.

Decision Taken:

- Chi-square = 24.87996158
- P value = 5.31843E-05
- p critical = 0.05
- p value < 0.05, Reject H_0
- chi-square > chi-square critical, Reject H_0

Managerial Inference:

H_0 is rejected, so there is a significant association between the level of exhaustion and mode of travel. This suggests that certain modes of travel are more exhausting, which could negatively impact students' ability to study or participate in other activities

chi-square observed				
Column1	Very	Ok	Not at all	Grand Total
Auto	8	26	34	68
Local Train	19	16	6	41
Uber/Ola	1	7	3	11
Grand Total	28	49	43	120

chi-square expected			
Column1	Very	Ok	Not at all
Auto	15.8666667	27.7666667	24.3666667
Local Train	9.56666667	16.7416667	14.6916667
Uber/Ola	2.56666667	4.49166667	3.94166667

chi-square calculation			
Column1	Very	Ok	Not at all
Auto	3.90028011	0.11240496	3.808527132
Local Train	9.3018583	0.03285631	5.142035356
Uber/Ola	0.95627706	1.40075758	0.224964764

MULTIPLE LINEAR REGRESSION

Objective of the Test: To examine CGPA depends on multiple independent factors such as distance, time, self-study time, travel commute, no of club joined, events participated, sleep hours, social media usage, and exhaustion.

Test Applied: Multiple linear regression.

Hypothesis Tested:

- **Null Hypothesis (H_0):** There is no significant difference between CGPA and different independent variable.(distance, time, self-study time, travel commute, no of club joined, events participated, sleep hours, social media usage, and exhaustion.)
- **Alternative Hypothesis (H_1):** There is a significant difference between CGPA and independent variables.

Calculations:

- Perform the regression using Excel's Data Analysis Toolpak.
- Obtain the coefficients and the p-value.

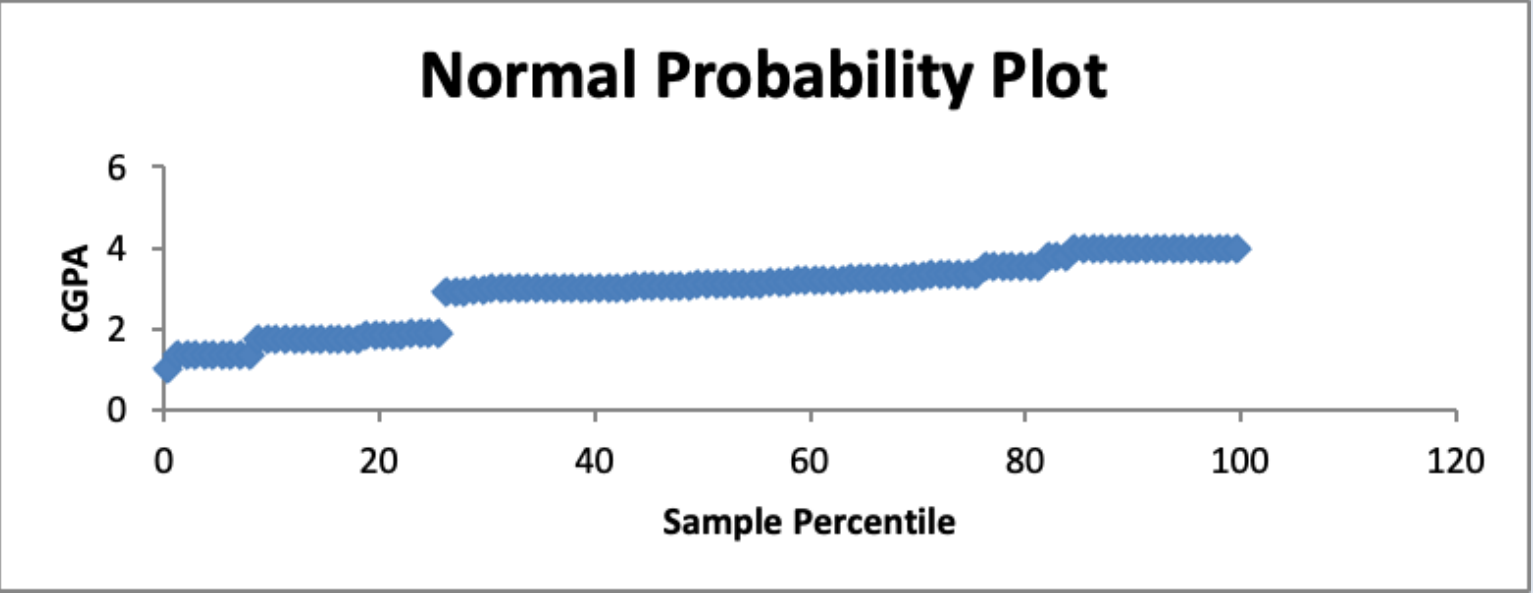
Managerial Inference:

- **Directly related Factors:** Commuting by auto and self-study time are positively associated with higher CGPA.
- **Inversely related Factors:** Distance (both ≤ 4 and > 4 units), commuting by local train, and social media usage impact CGPA.
- **Neutral Factors:** Participation in clubs, events, and reported levels of exhaustion do not show a significant impact on CGPA in this analysis.

MULTIPLE LINEAR REGRESSION

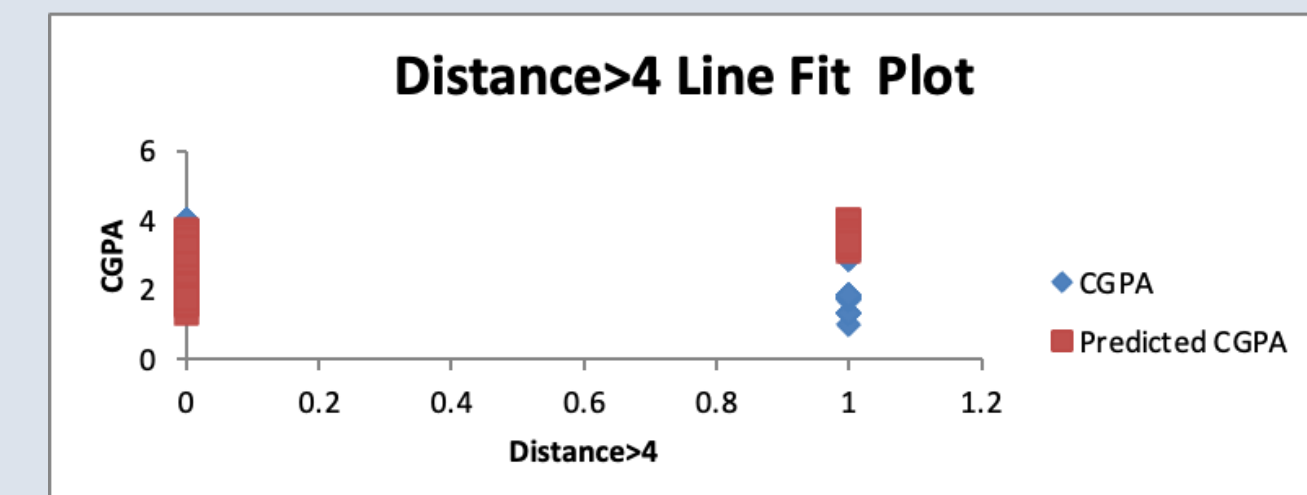
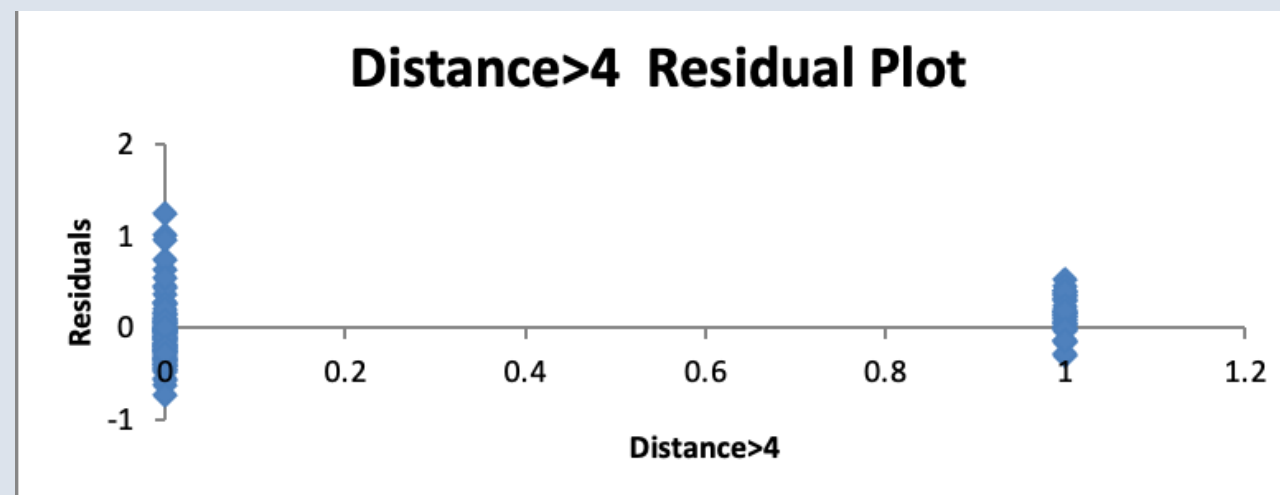
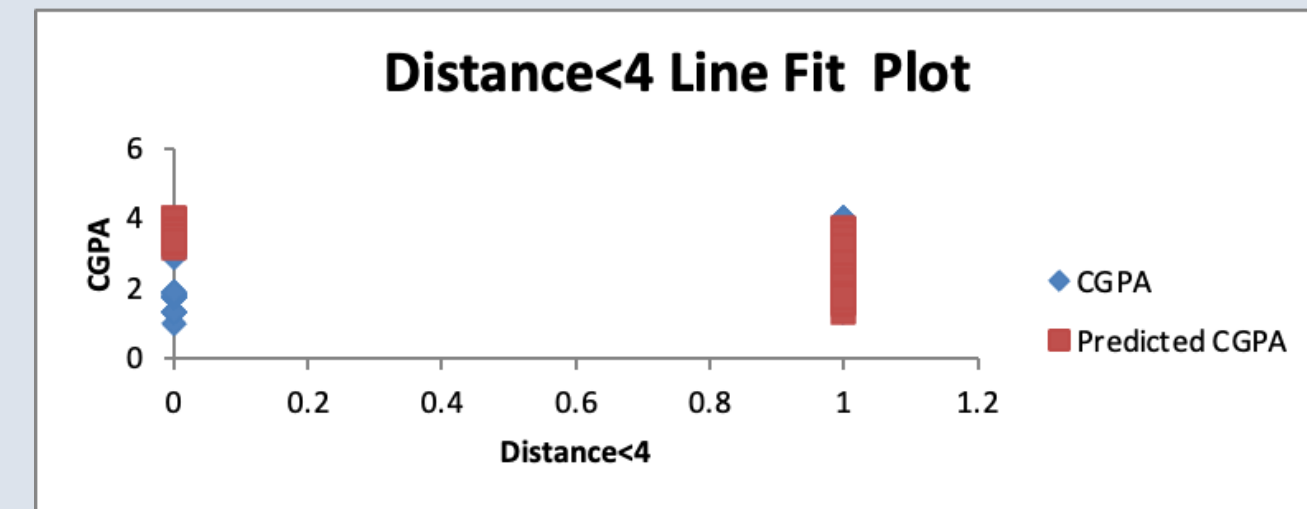
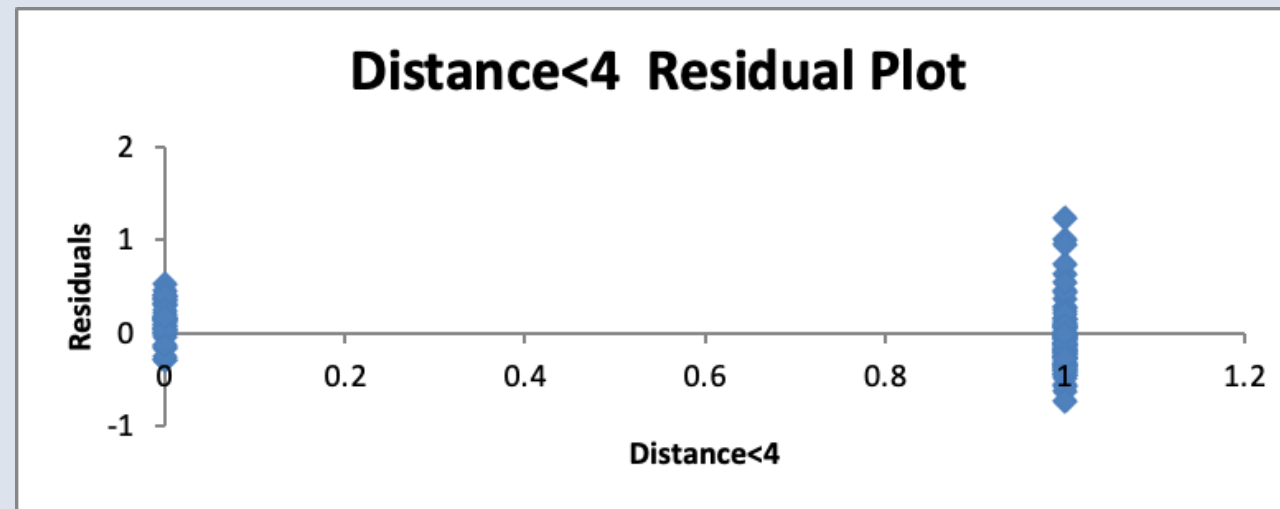
OUTPUT ANALYSIS

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.920079739								
R Square	0.846546727								
Adjusted R Square	0.827726986								
Standard Error	0.346165149								
Observations	120								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	13	70.0724538	5.390188754	44.98184768	4.28957E-37				
Residual	106	12.70201286	0.11983031						
Total	119	82.77446667							
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	2.095524814	0.459040985	4.565005918	1.35116E-05	1.185431414	3.005618214	1.185431414	3.005618214	
Distance<4	-0.956846415	0.411530141	-2.325094375	0.021971845	-1.772744945	-0.140947885	-1.772744945	-0.140947885	
Distance>4	-0.598767516	0.196386326	-2.030525873	0.044806809	-0.788122508	-0.009412524	-0.788122508	-0.009412524	
Travel Commute_A	1.547894925	0.374775751	4.130189644	7.25227E-05	0.804865548	2.290924302	0.804865548	2.290924302	
Travel Commute_It	-0.333360085	0.121395314	-2.746070449	0.007087472	-0.574038103	-0.092682068	-0.574038103	-0.092682068	
Travel Commute_u	-0.029647689	0.099664713	-0.297474277	0.766686652	-0.227242676	0.167947299	-0.227242676	0.167947299	
Self-study time	0.227154547	0.059897294	3.792400849	0.000248431	0.108402337	0.345906757	0.108402337	0.345906757	
Clubs Joined	0.052385327	0.033775573	1.550982626	0.123886489	-0.014578032	0.119348687	-0.014578032	0.119348687	
Events Participated	0.01365384	0.015231755	0.896406255	0.372067111	-0.016544595	0.043852275	-0.016544595	0.043852275	
Sleep Hours (hrs/da	0.034515761	0.029885361	1.154938737	0.250712464	-0.024734874	0.093766395	-0.024734874	0.093766395	
Exhaustion_NA	0.221907171	0.362770758	0.611700823	0.542045479	-0.49732114	0.941135483	-0.49732114	0.941135483	
Exhaustion_Ok	0.059117172	0.363555041	0.162608588	0.871136211	-0.661666056	0.7799004	-0.661666056	0.7799004	
Exhaustion_Very	0.185676843	0.366079983	0.507202938	0.613066029	-0.540112329	0.911466014	-0.540112329	0.911466014	
Social Media Usage	-0.049880932	0.019139401	-2.60619092	0.010472705	-0.087826655	-0.011935209	-0.087826655	-0.011935209	



MULTIPLE LINEAR REGRESSION

OUTPUT ANALYSIS

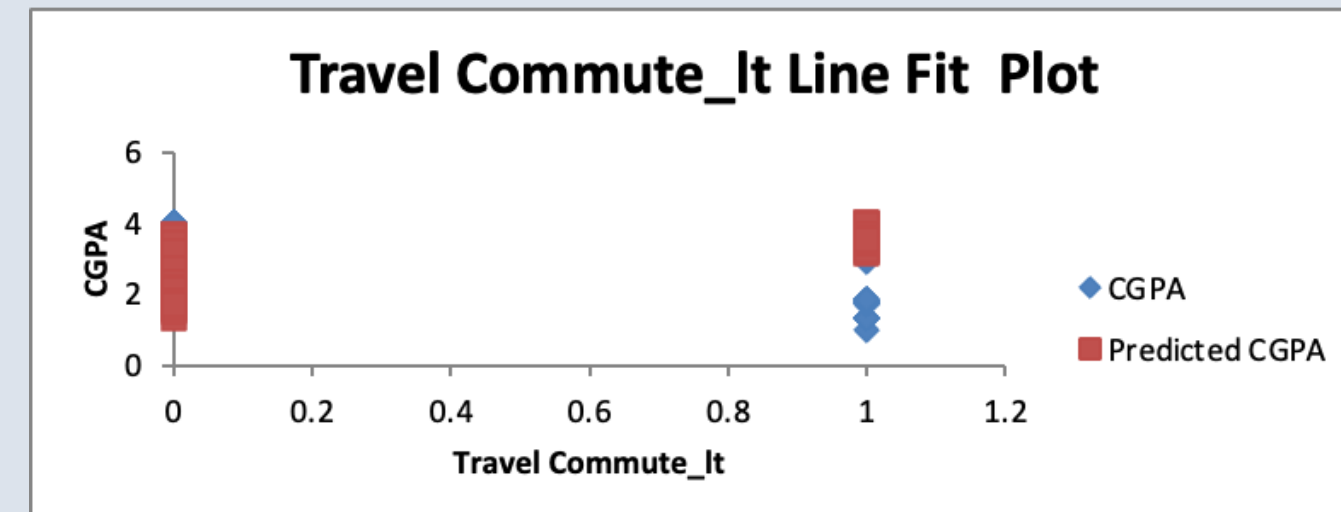
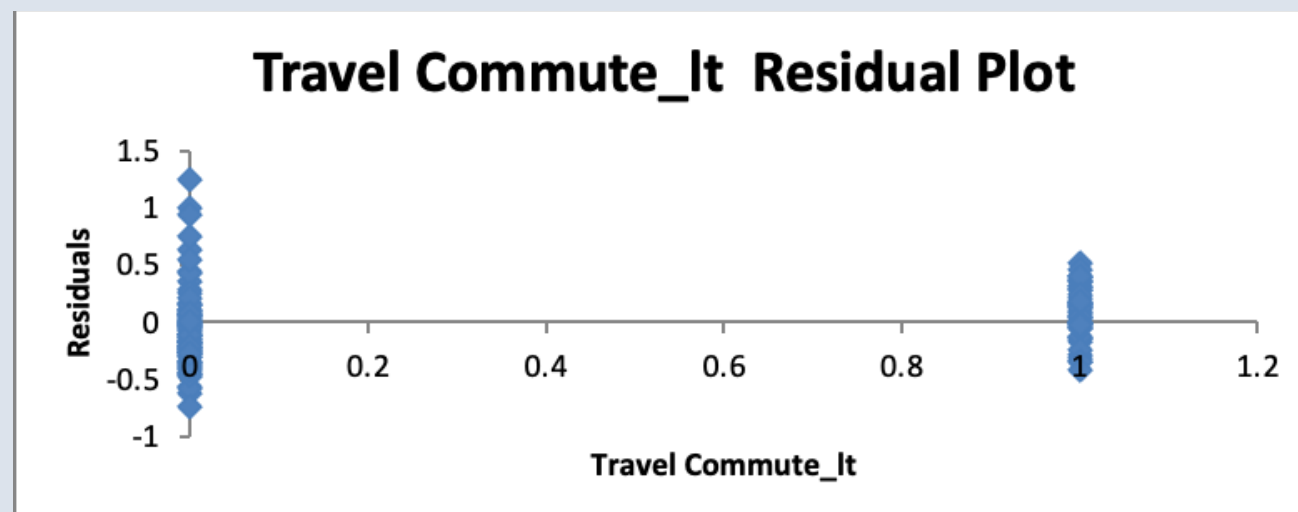
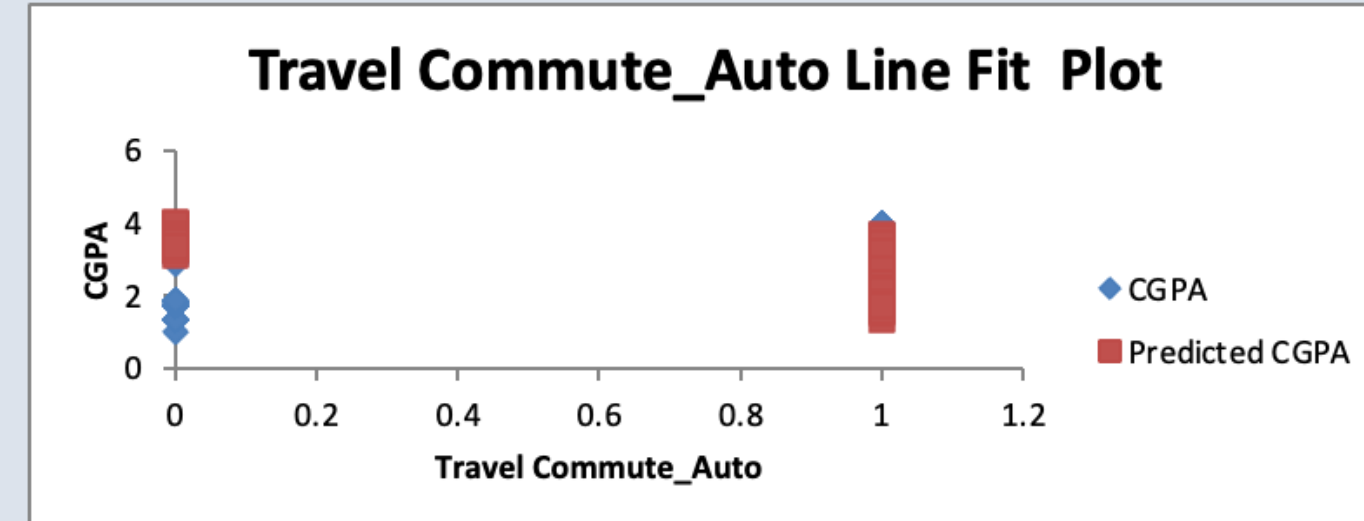
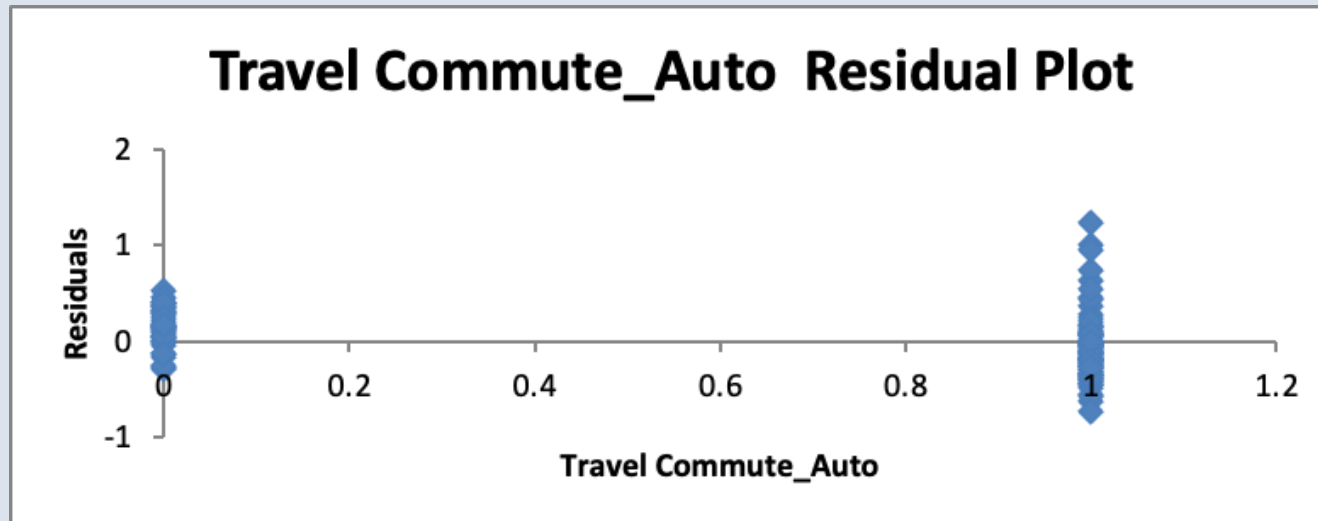


Managerial Inference:

- Distance≤4:
 - Coefficient: -0.9586
 - P-value: 0.0219 (significant at 5% level)
 - Inference: Students living within a distance of 4 units (miles/km, etc.) have a higher CGPA by 0.96 on average compared to those living farther away. This variable has a significant negative impact on CGPA.

MULTIPLE LINEAR REGRESSION

OUTPUT ANALYSIS

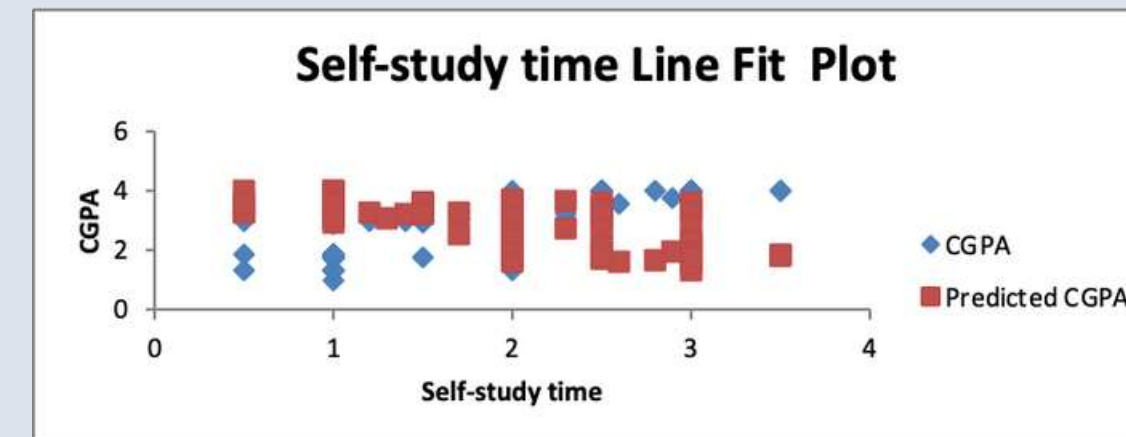
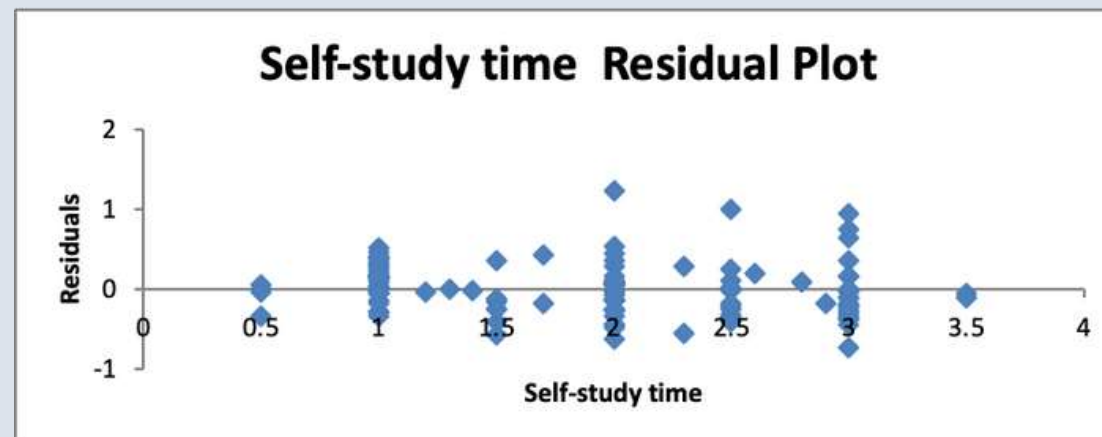


Managerial Inference:

- **Travel Commute_Auto:**
 - Coefficient: 1.54
 - P-value: 7.25e-05 (highly significant)
 - Inference: Students who commute by auto have a CGPA that is 1.54 points higher on average compared to those who do not commute by auto. This suggests that commuting by auto might be associated with factors that positively influence academic performance, such as flexibility in travel times or reduced stress from not relying on public transportation and also they could be living nearby to avail auto services.

MULTIPLE LINEAR REGRESSION

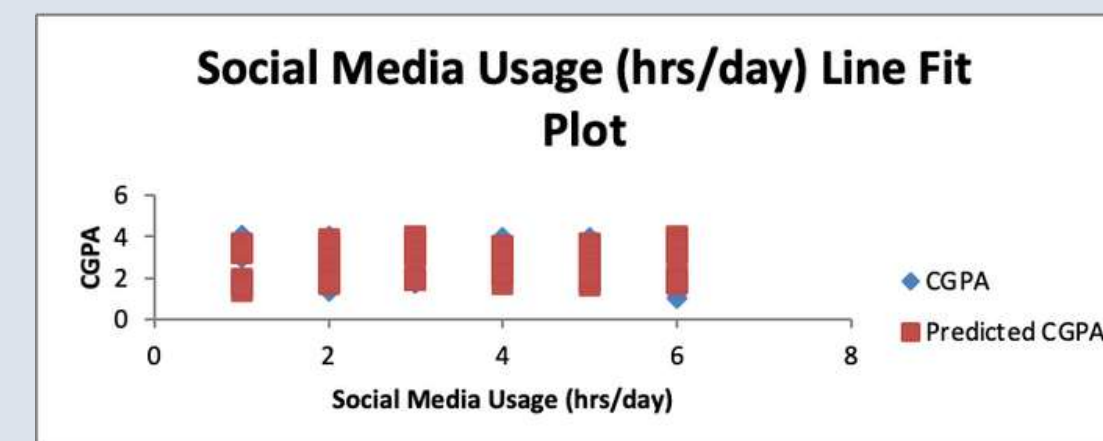
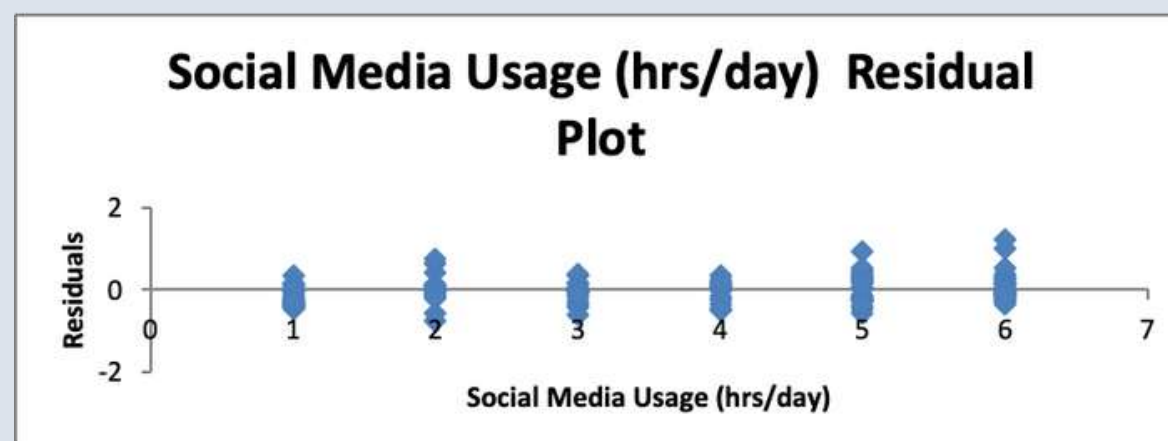
OUTPUT ANALYSIS



Managerial Inference:

Self-study time:

- Coefficient: 0.2275
- Inference: Every additional unit of self-study time increases CGPA by 0.23 points on average. This strong positive relationship highlights the importance of self-directed learning for academic success.



Managerial Inference:

Social Media Usage (hrs/day):

- Coefficient: -0.0499
- Inference: Increased social media usage is associated with a decrease in CGPA by 0.05 points on average. This relationship is statistically significant, suggesting that spending more time on social media could be detrimental to academic performance.



THANK YOU
