



Math Exams

Arham Raza | Ravleen Rattan | Russel Kwok | Jordan
Harjono | Jaiden Angeles - The Math Experts

“Being in a relationship distracts you from school”

-Mom

“Higher time socializing can lead to lower grades”

-ChatGPT

“C’S GET DEGREES”

-ALBERT EINSTEIN
(TRUST ME BRO)

“When I was your age, i had 4 jobs, took care of 3 cats and had a four hour walk to school”

-Dad

“The longer you study, the higher grades you’ll get”

-Fifth Grade Teacher

EXPERTS

P+



Ravleen



Russell



Arham



Jordan



Jaiden

P+
INTRODUCTION

THE DATA

ANALYSIS

CONCLUSION

THE SUMMARY

pt

Question

What factors lead to educational success?

Why Care

Considering students come from various backgrounds, we want to identify key factors that can lead to higher grades

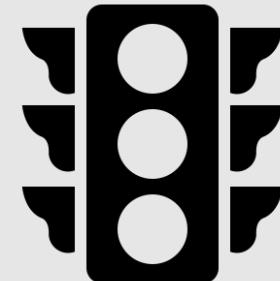
Solution

Methodology:

Linear Regression
Logistic Regression
CART

Areas Analyzed:

Relationships
Social Factors
School-Related Factors
Family Factors



pt

INTRODUCTION

THE DATA

ANALYSIS

CONCLUSION

WTD (WHAT THE DATA)

pt

Data Description



Student Performance Data Set

33 Variables, 396 Observations
Portugal-Based
High School is 3 Years
Grade Scale: 1 to 20
Below 10 is failing



Preliminary Screen

Variables Removed:
School, Guardian, Failures,
Paid, Nursery, Internet

Variable Categories

Social Factors:
Activities, Romantic,
Freetime, Goout, Dalc, Walc

School Factors:
Reason, Travelttime,
Studytime, Schoolsup,
Higher, Absences, G1, G2, G3

Personal Factors:
Sex, Age, Famsize, Address,
PStatus, Medu, Fedu, Mjob,
Fjob, Famsup, Famrel,
Health

INTRODUCTION

PT
THE DATA

ANALYSIS

CONCLUSION

WTD (WHAT THE DATA)

Limitations and Data Cleaning

pt

Limitations

- 1) The data is about two Portuguese high schools, forcing us to look at a smaller dataset
- 2) Only assessing Mathematics
- 3) G3 is highly dependent on G2

Data Cleaning - done through Excel

Removed students aged 19+

Converted character data to integers (ex: yes = 1, no = 0)

Removed final grades that were zero

Removed categorical variables

INTRODUCTION

PT
THE DATA

ANALYSIS

CONCLUSION

THE HYPOTHESES



Hypothesis 1: Does being in a relationship lead to a lower final grade ?

Hypothesis 2: Do highly social students have higher grades than less social students ?

Hypothesis 3: Do longer study times lead to higher final grades ?

Hypothesis 4: Do family factors lead to higher final grades ?

Hypothesis 5: Do student surroundings affect their higher education decision ?

METHODOLOGY

Analysis Methods

pt

MODELS

- Multiple Linear Regression with *G3* as the dependent variable
- Logistic Regression with *G3* as the dependent variable (Pass/Fail)
- Logistic Regression with *Higher* as the dependent variable

METHODS

- Create multiple models with *G3* and *Higher* as the dependent variable
- Use kitchen sink model and adjust IV's
- Use stepwise to improve model
- Compare and choose the best model

INTRODUCTION

THE DATA

P⁺
ANALYSIS

CONCLUSION

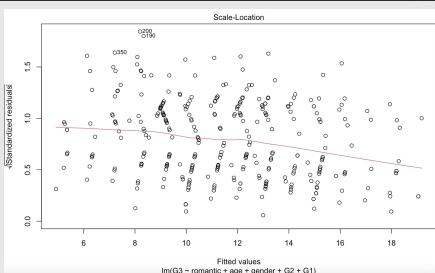
HYPOTHESIS I

pt

	Dependent variable:		
	(1)	G3 (2)	(3)
romantic	-0.103 (0.097)	-0.103 (0.098)	-0.108 (0.294)
age	0.015 (0.040)	0.015 (0.040)	0.023 (0.041)
Gender		-0.0005 (0.089)	0.013 (0.090)
G2	0.883*** (0.035)	0.883*** (0.035)	0.882*** (0.035)
G1	0.114*** (0.033)	0.114*** (0.033)	0.111*** (0.033)
goout			-0.084* (0.050)
romantic:goout			0.002 (0.088)
Constant	0.002 (0.705)	0.003 (0.707)	0.169 (0.711)
Observations	352	352	352
R2	0.934	0.934	0.935
Adjusted R2	0.933	0.933	0.933
Residual Std. Error	0.830 (df = 347)	0.831 (df = 346)	0.828 (df = 344)
F Statistic	1,226.295*** (df = 4; 347)	978.209*** (df = 5; 346)	703.510*** (df = 7; 344)

Note:

*p<0.1; **p<0.05; ***p<0.01



INTRODUCTION

Does Being In A Relationship Lead to Lower Grades : ?

Model 1:
Highest Fstat
No significant relevant variables



AIC:
Model 1: 874.4817
Model 2: 876.4816
Model 3: 874.3292

The coefficient for *romantic* is not significant

There is a negative relationship between *romantic* and *G3*, however, we cannot reject the null hypothesis

Conclusion: Students in relationships do not necessarily get a lower final grade than students without relationships



pt
ANALYSIS

THE DATA

CONCLUSION

HYPOTHESIS 2

pt

?

Does Being Highly Social Lead to Higher Grades :

	Dependent variable:				
	(1)	(2)	G3 (3)	(4)	(5)
goout	-0.537*** (0.161)		-0.365** (0.172)	-0.381** (0.178)	-0.387** (0.178)
Walc		-0.415** (0.171)	-0.277 (0.182)	-0.267 (0.183)	-0.269 (0.183)
Dalc			-0.043 (0.244)	-0.077 (0.247)	-0.063 (0.248)
freetime	0.078 (0.176)	-0.011 (0.173)		0.106 (0.177)	0.087 (0.178)
activities	0.354 (0.340)	0.299 (0.341)	0.343 (0.337)		0.325 (0.340)
Constant	12.754*** (0.671)	12.425*** (0.631)	13.184*** (0.551)	13.096*** (0.670)	12.990*** (0.680)
Observations	352	352	352	352	352
R2	0.034	0.032	0.045	0.043	0.045
Adjusted R2	0.025	0.021	0.034	0.032	0.031
Residual Std. Error	3.169 (df = 348)	3.176 (df = 347)	3.155 (df = 347)	3.158 (df = 347)	3.159 (df = 346)
F Statistic	4.050*** (df = 3; 348)	2.887** (df = 4; 347)	4.049*** (df = 4; 347)	3.873*** (df = 4; 347)	3.281*** (df = 5; 346)
Note:	*p<0.1; **p<0.05; ***p<0.01				

Model 3:
Highest Fstat
Highest adjusted R2
 $p<0.01$

AIC:
Model 1: 1816.59
Model 2: 1816.81
Model 3: 1814.83
Model 4: 1815.52
Model 5: 1816.59

Relevant variables are not significant except goout

Model 3 has the highest adjusted R2 but is still very small

Conclusion: students who spend more time on social activities do not necessarily get better or worse grades



HYPOTHESIS 3

pt

	Dependent variable:		
	(1)	G3 (2)	(3)
studytime	0.494** (0.199)	0.039 (0.055)	0.103 (0.449)
schoolsup	-2.239*** (0.473)	-0.187 (0.132)	-2.235*** (0.471)
G2		0.982*** (0.015)	
traveltime			-0.913 (0.586)
studytime:traveltime			0.242 (0.277)
Constant	10.832*** (0.444)	0.316 (0.200)	12.244*** (0.986)
Observations	352	352	352
R2	0.075	0.932	0.086
Adjusted R2	0.070	0.931	0.075
Residual Std. Error	3.096 (df = 349)	0.842 (df = 348)	3.086 (df = 347)
F Statistic	14.121*** (df = 2; 349)	1,582.557*** (df = 3; 348)	8.157*** (df = 4; 347)
Note:	*p<0.1; **p<0.05; ***p<0.01		



INTRODUCTION

THE DATA

pt
ANALYSIS

CONCLUSION

Task: Does Longer Study Time lead to Higher Grades? ?

Model 2:
Highest Fstat
Highest Adjusted R2



AIC:
Model 1: 1799.502
Model 2: 884.1434
Model 3: 1799.259

The coefficient for *studytime* and *schoolsup* is not significant

There is a positive relationship between *studytime* and *G3*, however, we cannot reject the null hypothesis

Conclusion: Students that study longer do not necessarily get a higher grade



HYPOTHESIS 4

pt

Task: Do Family Factors Affect Student Grades



Model 3:
Highest Fstat
 $p < 0.01$
All variables have $p < 0.1$



AIC:
Model 1: 868.0427
Model 2: 874.2018
Model 3: 874.813

Holding Rest of the Model Constant:

Higher Mother Education → Lower Grade
Higher Father Education → Lower Grade



Higher Mother/Father Education → Higher Grade
Better Family Relationship → Higher Grade



	Dependent variable:		
	G3 (1)	(G3) (2)	G3 (3)
FatherEducation	-0.055 (0.054)		-0.293** (0.132)
MotherEducation	-0.052 (0.057)		-0.191* (0.110)
MotherJob	-0.081** (0.037)		
FatherJob	0.009 (0.037)		
FamSupport	0.043 (0.093)		
PStatus	0.169 (0.144)		
FamRel	0.162*** (0.050)	0.162*** (0.050)	0.165*** (0.050)
MotherEducation:FatherEducation			0.084** (0.042)
G2	0.894*** (0.032)	0.989*** (0.014)	0.992*** (0.014)
G1	0.104*** (0.032)		
Constant	-0.131 (0.397)	-0.343 (0.260)	0.227 (0.366)
Observations	352	352	352
R2	0.937	0.933	0.934
Adjusted R2	0.935	0.933	0.933
Residual Std. Error	0.816 (df = 342)	0.832 (df = 349)	0.829 (df = 346)
F Statistic	564.665*** (df = 9; 342)	2,438.959*** (df = 2; 349)	983.186*** (df = 5; 346)
Note:	$*p < 0.1$; ** $p < 0.05$; *** $p < 0.01$		

HYPOTHESIS 5

pt

	Dependent variable:		
	(1)	higher (2)	(3)
Medu	0.824** (0.345)	1.371* (0.703)	1.132*** (0.429)
Fedu	0.405 (0.376)	0.225 (0.592)	
travelttime		-0.756 (0.559)	-0.755* (0.424)
studytime		1.043 (0.892)	1.014 (0.682)
failures		-0.238 (0.556)	
famsup	0.372 (0.623)	0.466 (1.185)	
paid		2.564* (1.502)	2.317* (1.196)
internet		1.915 (1.229)	2.005** (0.934)
absences		-0.129** (0.054)	-0.099*** (0.034)
G3		-0.007 (0.643)	
schoolsup	0.313 (1.101)	-0.390 (1.675)	
address	-0.114 (0.651)	2.021 (1.561)	1.993* (1.031)
Gender	-1.435** (0.705)	-2.669** (1.304)	-2.344** (1.004)
age		-0.921* (0.504)	-0.819** (0.344)
Constant	1.389 (1.258)	13.157 (10.945)	12.375** (5.997)
Observations	352	352	352
Log Likelihood	-44.912	-26.976	-31.732
Akaike Inf. Crit.	103.825	115.951	83.465

Note: *p<0.1; **p<0.05; ***p<0.01

Task: Do student support factors affect their higher education decision.



Model 3:

Lower AIC

Lower Log Likelihood

All Variables have p<0.1



AIC:

Model 1: 103.825

Model 2: 115.951

Model 3: 83.465

Living Urban → higher education
Higher Mother's Education → higher education



Age → decreases higher education



Conclusion: Student support factors do affect their higher education decision

DATA PROBLEM: HIGHER EDUCATION



Imbalance data 95% of students pursued higher education

Overemphasis on majority

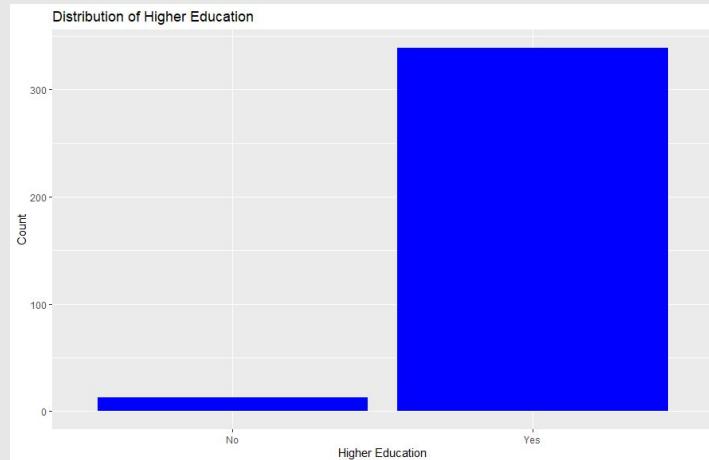


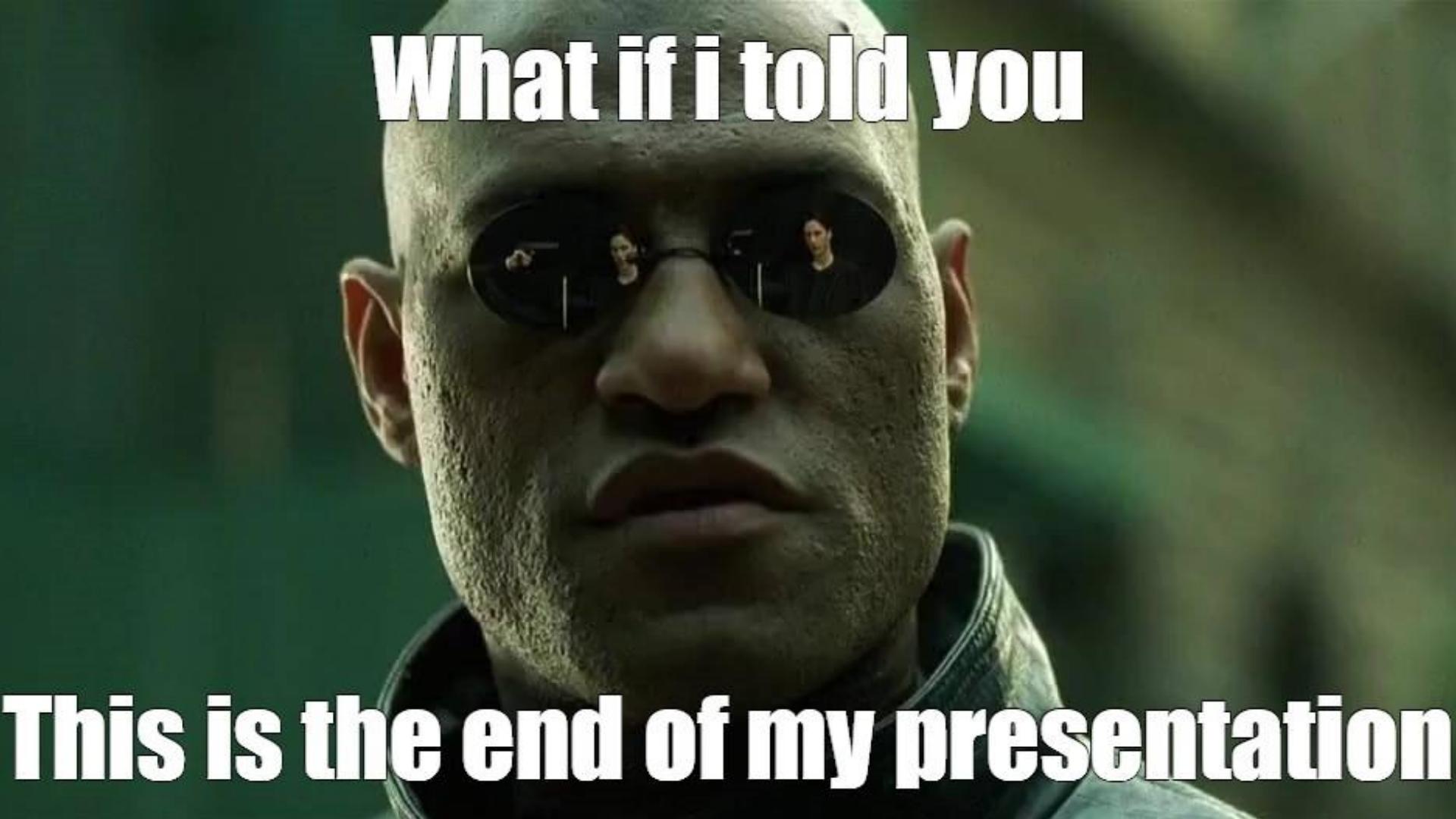
Less nuanced understanding

Minimal interactions with minority class



Lack of understanding between independent variables and majority





What if i told you

This is the end of my presentation

LET'S LOOK AT IT ALL TOGETHER

pt

Task: What overall factors lead to the highest grades?

Similar approach as previous models

Holding Rest of the Model Constant:

Higher Current Health → Lower Grade
Higher amount of Going Out → Lower Grade



Higher G2 → Higher Grade
Higher G1 → Higher Grade
Higher Family Relationship → Higher Grade



Dependent variable:	
	G3
G2	0.869*** (0.033)
G1	0.116*** (0.031)
absences	-0.011* (0.005)
health	-0.066** (0.031)
goout	-0.084** (0.040)
famrel	0.164*** (0.049)
Constant	0.263 (0.316)
<hr/>	
Observations	352
R2	0.938
Adjusted R2	0.937
Residual Std. Error	0.807 (df = 345)
F Statistic	867.755*** (df = 6; 345)
<hr/>	
Note:	*p<0.1; **p<0.05; ***p<0.01

THERE'S MORE?



Task: What if a student just wants to pass?

Similar approach as previous models

Holding Rest of the Model Constant:

Higher Current Health → Lower log odds of passing
Higher amount of Going Out → Lower log odds of passing



Higher G2 → Higher log odds of passing
Higher G1 → Higher log odds of passing
Higher Family Relationship → higher log odds of passing



Dependent variable:	
pass	
G2	2.313*** (0.364)
absences	-0.059** (0.027)
health	-0.366* (0.207)
goout	-0.464** (0.236)
famrel	1.020*** (0.370)
studytime	-0.693* (0.356)
traveltime	0.700** (0.313)
Constant	-21.406*** (3.958)
<hr/>	
Observations	352
Log Likelihood	-53.494
Akaike Inf. Crit.	122.988
<hr/>	
Note:	*p<0.1; **p<0.05; ***p<0.01

METHODOLOGY

pt

Prediction Method - CART

MODELS

- CART model → Classification Tree

METHODS

- Transformed G3 into a binary variable
 - Grade > 10 → Pass → 1
 - Grade < 10 → Fail → 0
- Created Training and Testing Data (80:20)
- Variables Used: G3, G2, Absences, Famrel, Studytime, Health, Goout, Travelttime

INTRODUCTION

THE DATA

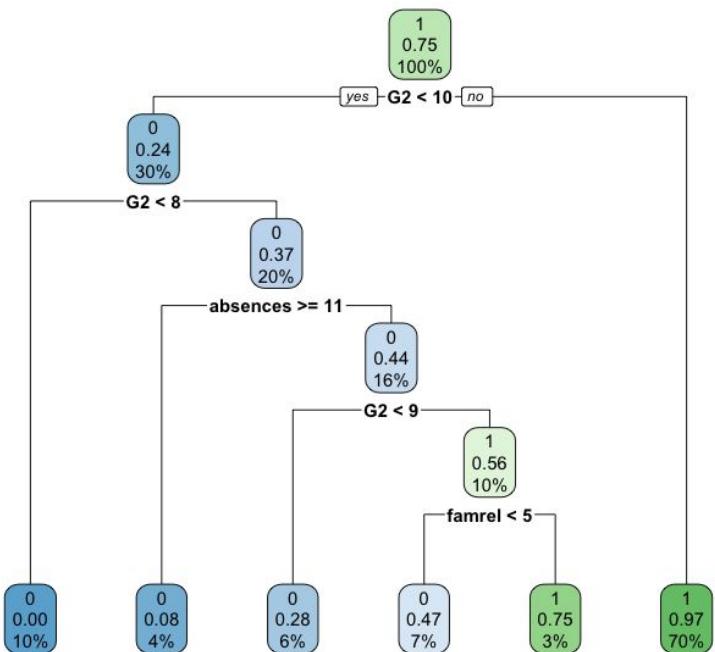
ANALYSIS

CONCLUSION

CART MODEL

pt

Result



Two CART Models Tested:

80/20 Split

70/30 Split

Confusion Matrix (Testing Data) "1" → Pass "0" → Fail		
Prediction	Fail	Pass
Actual (Fail)	16	2
Actual (Pass)	1	51
Accuracy: 95% Recall: 89% Precision: 94% F1: 0.91		

INTRODUCTION

THE DATA

pt
ANALYSIS

CONCLUSION

MOST IMPORTANT VARIABLES



LOGIT

CART

G2

Absences

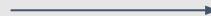
Health

Goout

Famrel

Studytime

Traveltim



G2

Absences

Famrel

Key Findings

KEY FINDINGS



G2 (second period grade) has a great influence on the final grade



Good quality of family relations has a positive impact on the final grade



Increased travel time leads to a higher final grade



Higher study time leads to a lower grade

Good health status leads to a lower final grade



Various Support factors lead to pursuing higher education



Higher absenteeism leads to lower final grade

NEXT STEPS



FURTHER ANALYSIS

- Analyze study time, health status, and travel time further

CONSIDER G1 AND G2 AS DEPENDENT VARIABLES

- Find the impact of the variables on first and second semester grades

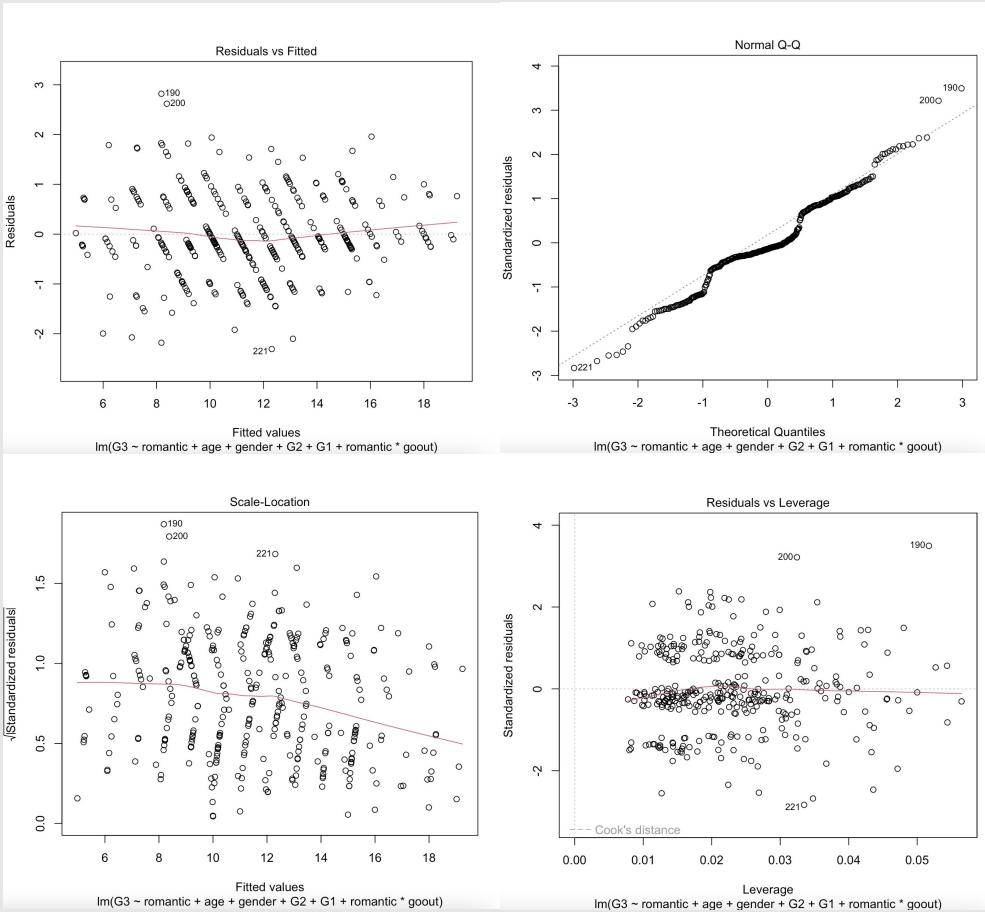
LARGER DATASET

- Compare our analysis with a larger dataset to see if there are similar results

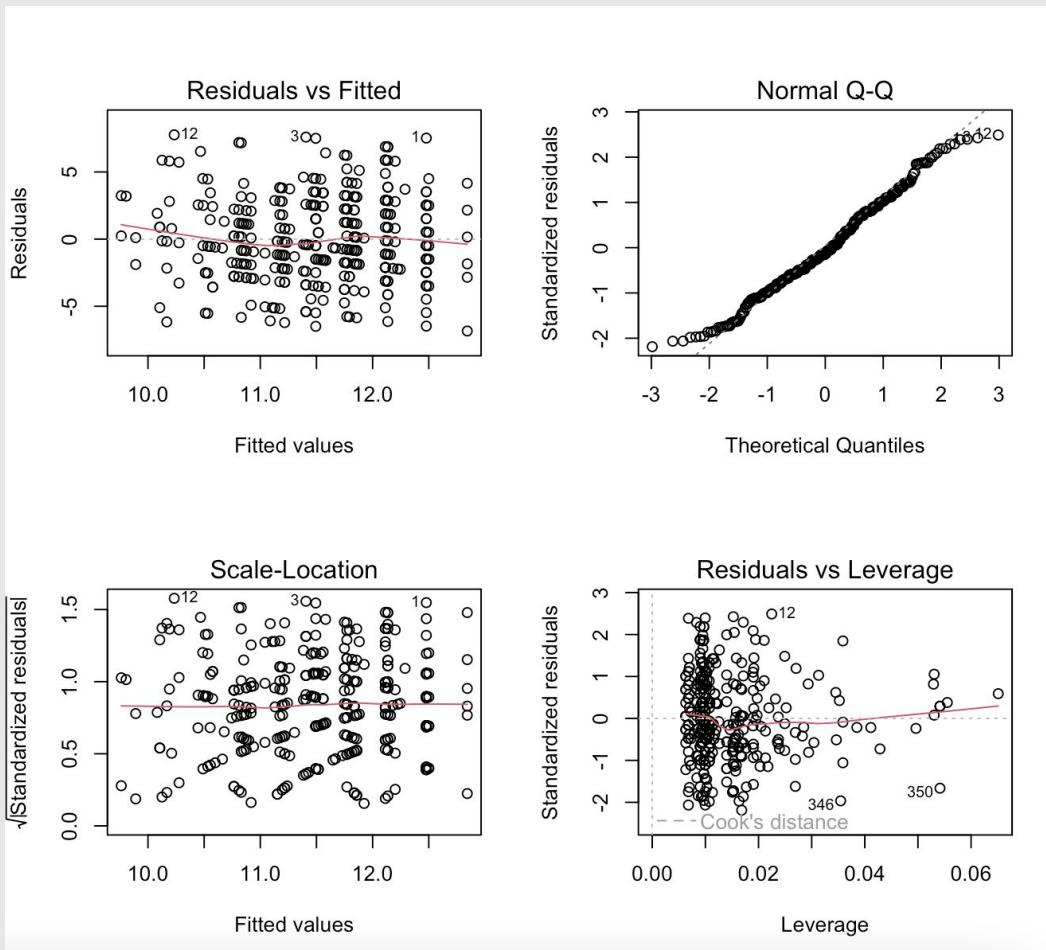


Thank You For
Listening

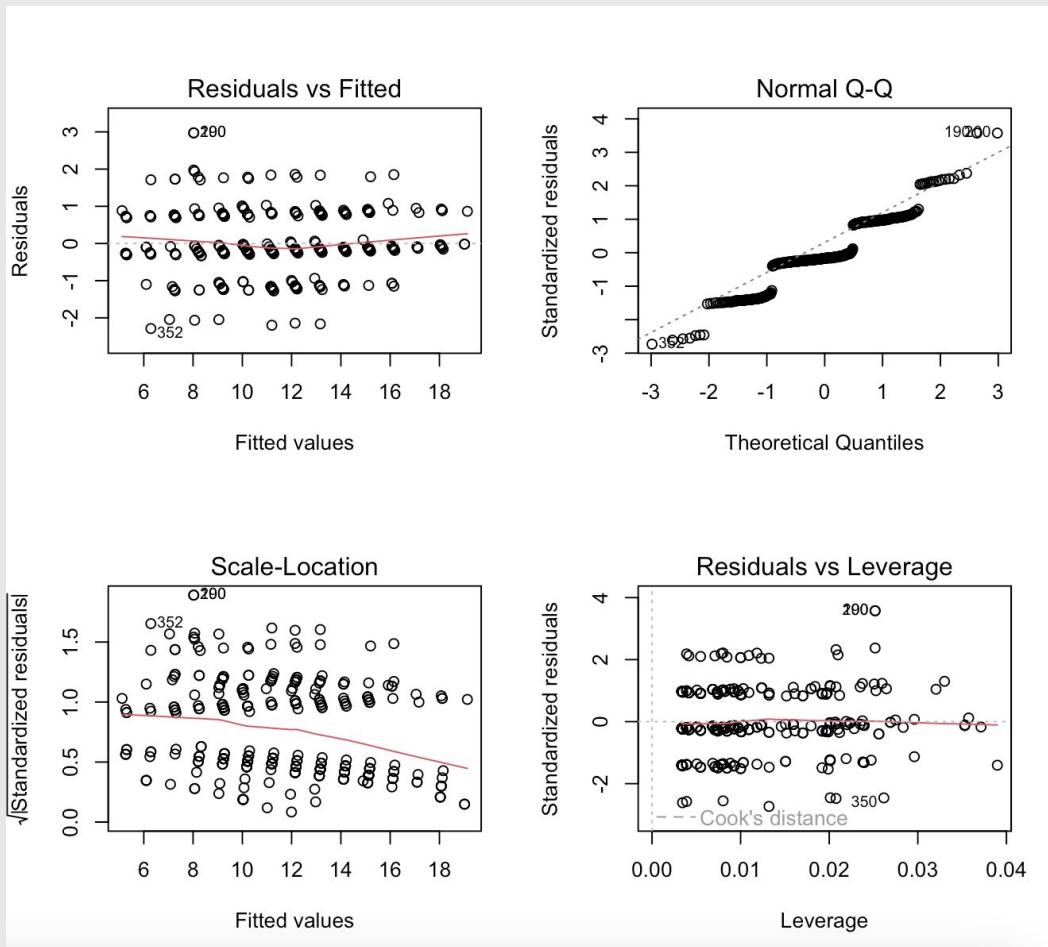
Appendix 1: Hypothesis 1 model's diagnostics



Appendix 2: Hypothesis 2 model's diagnostics

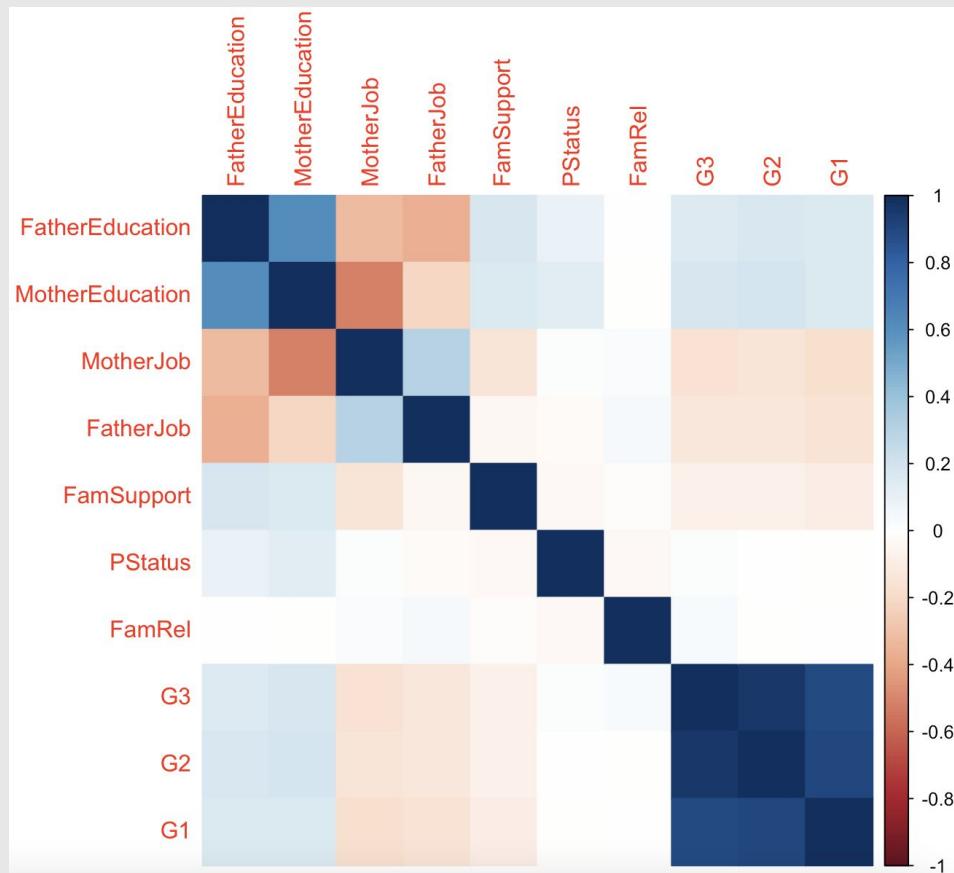


Appendix 3: Hypothesis 3 model's diagnostics

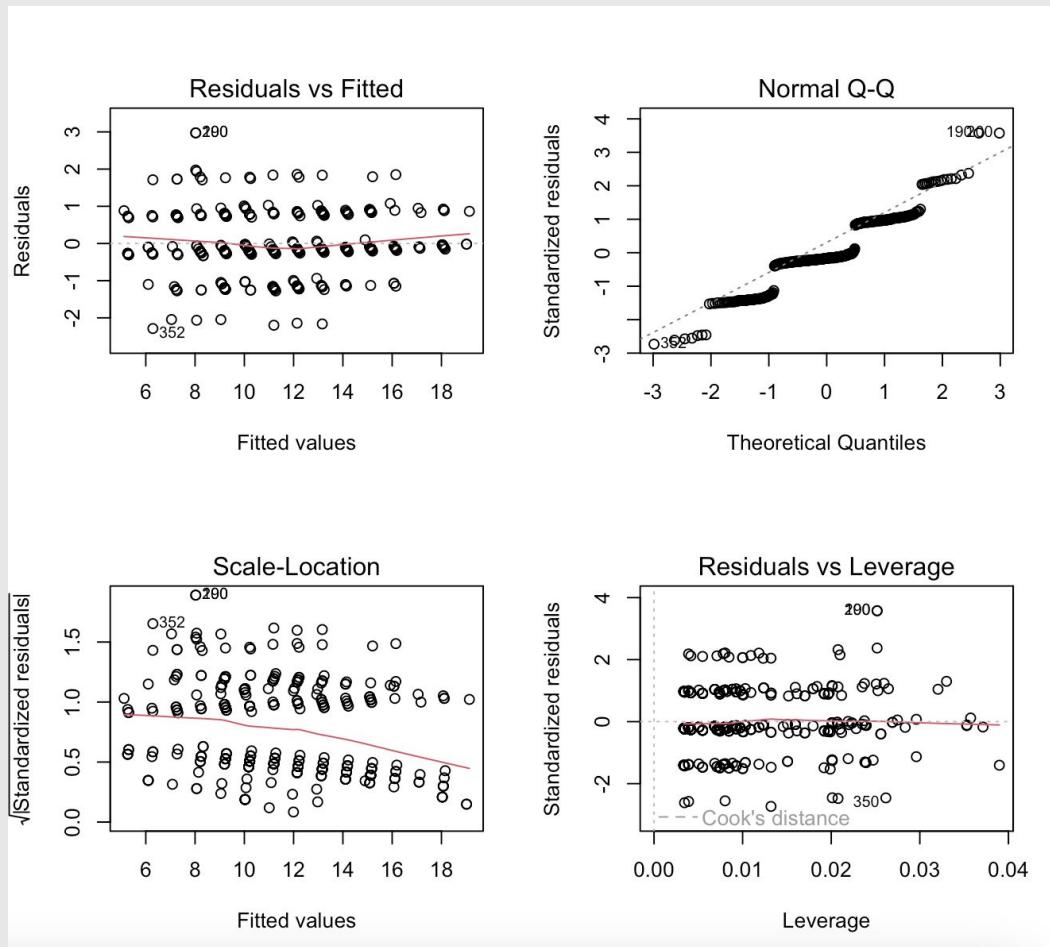


Appendix 4: Hypothesis 4 correlation

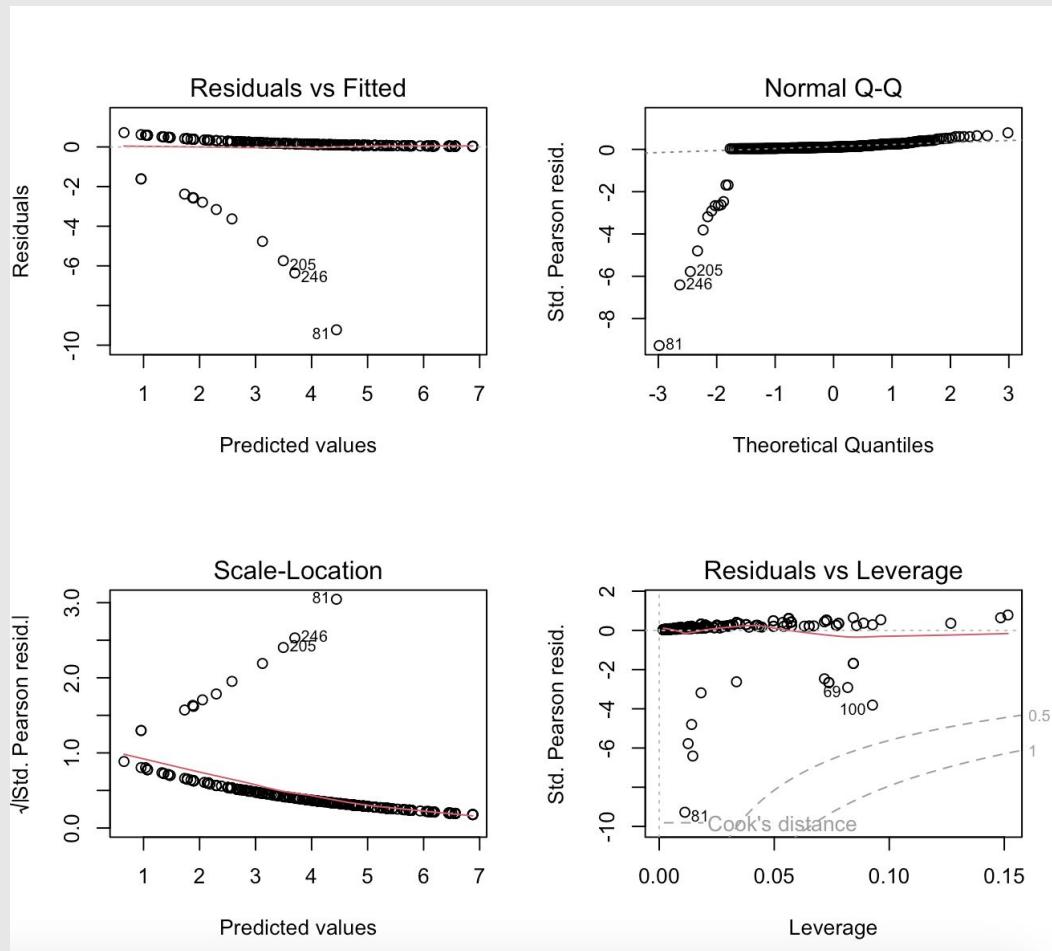
Correlation Matrix Family Factors



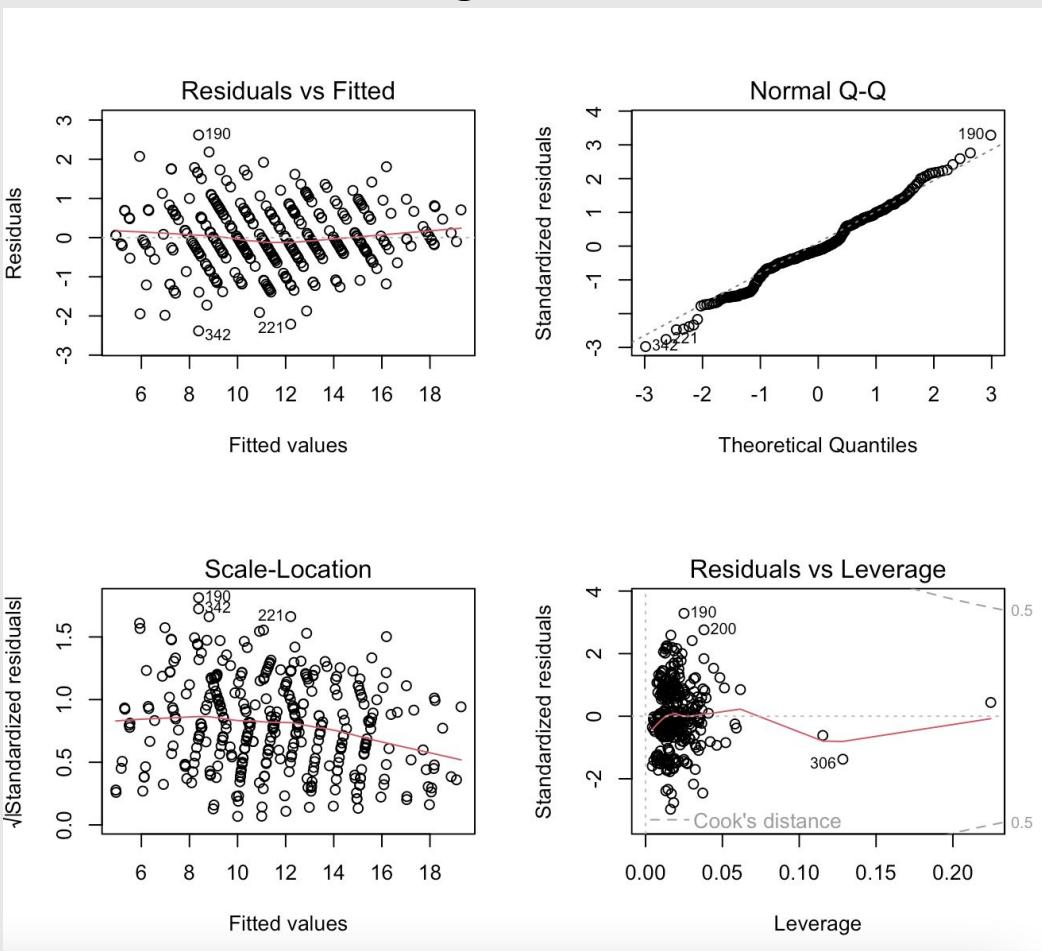
Appendix 5: Hypothesis 4 model's diagnostics



Appendix 6: Hypothesis 5 model's diagnostics



Appendix 7: Hypothesis 6 model's diagnostics

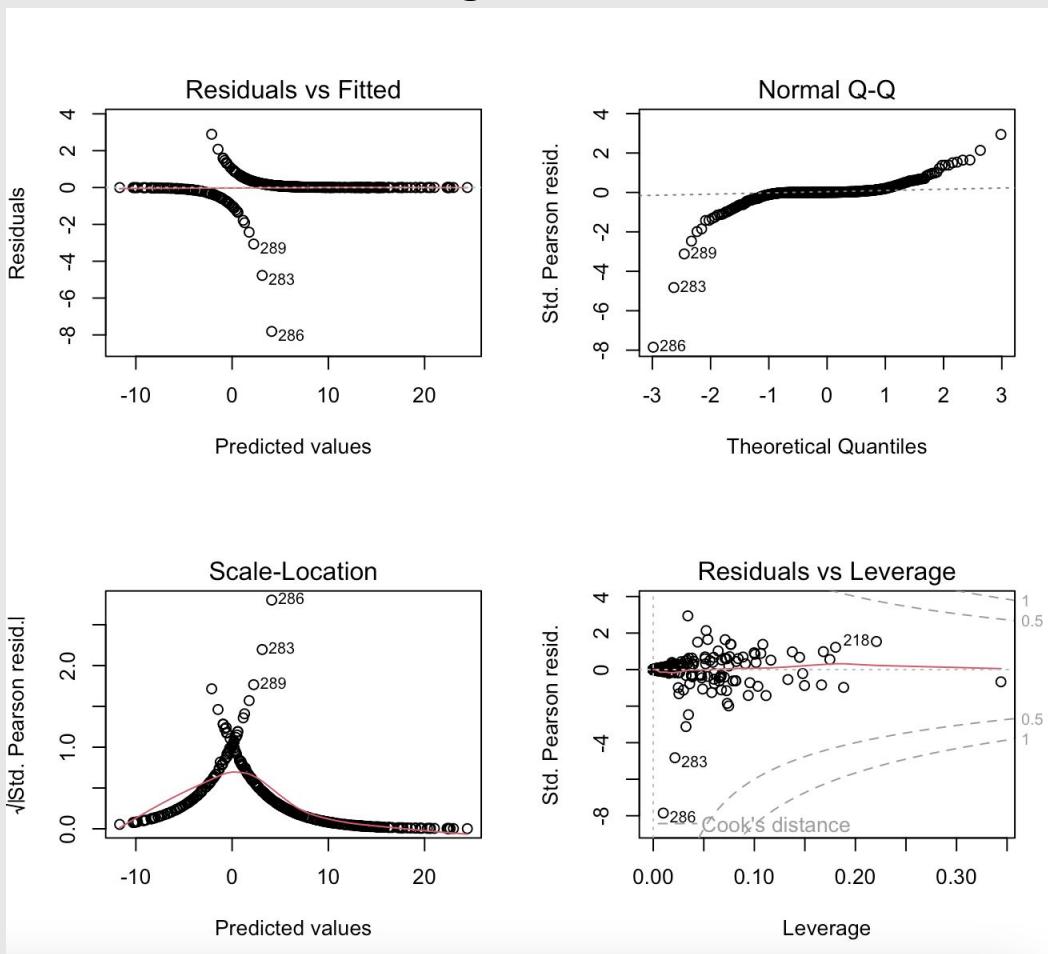


Appendix 8: Hypothesis 6 stargazer

	Dependent variable: G3		
	(1)	(2)	(3)
G2	0.880*** (0.035)	0.864*** (0.033)	0.869*** (0.033)
G1	0.102*** (0.034)	0.120*** (0.031)	0.116*** (0.031)
absences	-0.013** (0.006)	-0.011** (0.005)	-0.011* (0.005)
health	-0.072** (0.032)	-0.068** (0.031)	-0.066** (0.031)
goout	-0.094* (0.048)	-0.099** (0.044)	-0.084** (0.040)
freetime	-0.016 (0.049)		
famrel	0.177*** (0.052)	0.166*** (0.050)	0.164*** (0.049)
romantic	-0.021 (0.099)		
higher	-0.173 (0.250)		
activities	-0.015 (0.093)		
famsup	0.057 (0.096)		
schoolsup	-0.129 (0.140)		
studytime	0.001 (0.060)		
traveltime	0.021 (0.070)		

reason	0.050 (0.048)		
fjob	0.005 (0.038)		
mjob	-0.082** (0.039)		
fedu	-0.046 (0.055)		
medu	-0.038 (0.060)		
pstatus	0.196 (0.147)		
address	-0.126 (0.116)		
age	0.022 (0.044)		
gender	-0.019 (0.102)		
walc	0.013 (0.051)	-0.069 (0.071)	
dalc	0.014 (0.067)	-0.345* (0.197)	
walc:dalc		0.082* (0.045)	
Constant	0.607 (0.952)	0.649 (0.397)	0.263 (0.316)
Observations	352	352	352
R2	0.940	0.939	0.938
Adjusted R2	0.936	0.937	0.937
Residual Std. Error	0.815 (df = 326)	0.806 (df = 342)	0.807 (df = 345)
F Statistic	204.954*** (df = 25; 326)	579.933*** (df = 9; 342)	867.755*** (df = 6; 345)
Note:	*p<0.1; **p<0.05; ***p<0.01		

Appendix 9: Hypothesis 7 model's diagnostics



Appendix 10: Hypothesis 7 stargazer

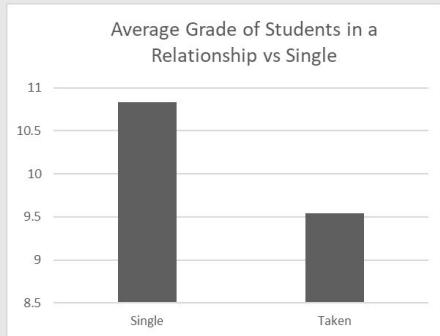
Dependent variable:			
	pass		
	(1)	(2)	(3)
G2	2.313*** (0.364)	2.164*** (0.382)	2.381*** (0.373)
G1		0.331* (0.194)	
absences	-0.059** (0.027)	-0.074** (0.031)	-0.066** (0.029)
health	-0.366* (0.287)	-0.363* (0.212)	-0.385* (0.215)
goout	-0.464** (0.236)	-0.407 (0.250)	0.129 (0.577)
famrel	1.020*** (0.370)	0.945** (0.380)	1.068*** (0.387)
studytime	-0.693* (0.356)	-1.330* (0.768)	0.225 (0.958)
traveltime	0.700** (0.313)	0.028 (0.806)	0.546* (0.332)
fjob		0.486** (0.229)	0.447** (0.227)
studytime:traveltime		0.311 (0.401)	
goout:studytime			-0.308 (0.286)
Constant	-21.406*** (3.958)	-23.416*** (4.304)	-25.463*** (4.808)
<hr/>			
Observations	352	352	352
Log Likelihood	-53.494	-49.654	-50.787
Akaike Inf. Crit.	122.988	121.309	121.574
<hr/>			
Note:	*p<0.1; **p<0.05; ***p<0.01		

Appendix 11: Summary stats of relevant variables

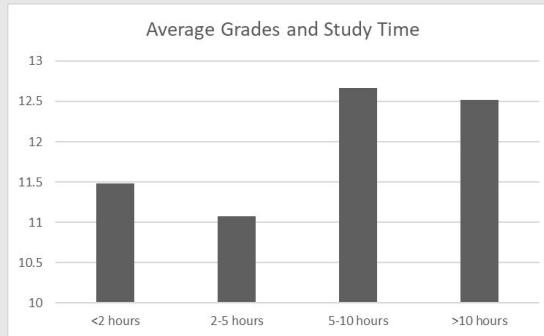
```
> summary(test)
   Gender      age      address     Pstatus      Medu      Fedu      Mjob      Fjob      reason      traveltime      studytime      schoolsup
Gender Min. :0.0000  Min. :15.0  Min. :1.000  Min. :1.000  Min. :0.000  Min. :0.00  Min. :1.000  Min. :1.000  Min. :1.000  Min. :1.000  Min. :1.000  Min. :0.000
1st Qu.:0.0000  1st Qu.:16.0  1st Qu.:1.000  1st Qu.:1.000  1st Qu.:2.000  1st Qu.:2.00  1st Qu.:3.000  1st Qu.:3.000  1st Qu.:1.000  1st Qu.:1.000  1st Qu.:1.000  1st Qu.:1.000  1st Qu.:0.000
Median :0.0000  Median :17.0  Median :1.000  Median :1.000  Median :3.000  Median :3.00  Median :3.500  Median :5.000  Median :2.000  Median :1.000  Median :2.000  Median :1.000  Median :0.000
Mean   :0.4773  Mean   :16.6  Mean   :1.219  Mean   :1.105  Mean   :2.798  Mean   :2.56  Mean   :3.463  Mean   :3.946  Mean   :2.264  Mean   :1.435  Mean   :2.048  Mean   :0.142
3rd Qu.:1.0000  3rd Qu.:18.0  3rd Qu.:1.000  3rd Qu.:1.000  3rd Qu.:4.000  3rd Qu.:3.25  3rd Qu.:5.000  3rd Qu.:5.000  3rd Qu.:3.000  3rd Qu.:2.000  3rd Qu.:2.000  3rd Qu.:0.000  Max.   :1.000
Max.   :1.0000  Max.   :19.0  Max.   :2.000  Max.   :2.000  Max.   :4.000  Max.   :4.00  Max.   :5.000  Max.   :5.000  Max.   :4.000  Max.   :4.000  Max.   :4.000  Max.   :1.000
famsup      activities      higher      romantic      famrel      freetime      goout      dalc      walc      health      absences
famsup Min. :0.0000  Min. :0.0000  Min. :0.0000  Min. :0.0000  Min. :1.00  Min. :1.000  Min. :1.000  Min. :1.000  Min. :1.000  Min. :1.000  Min. :0.00
1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:1.0000  1st Qu.:0.0000  1st Qu.:4.00  1st Qu.:3.000  1st Qu.:2.000  1st Qu.:1.000  1st Qu.:1.000  1st Qu.:3.000  1st Qu.:1.000  1st Qu.:2.00
Median :1.0000  Median :1.0000  Median :1.0000  Median :0.0000  Median :4.00  Median :3.000  Median :3.000  Median :1.000  Median :2.000  Median :4.000  Median :4.000  Median :4.00
Mean   :0.6165  Mean   :0.5085  Mean   :0.9631  Mean   :0.3125  Mean   :3.94  Mean   :3.227  Mean   :3.091  Mean   :1.477  Mean   :2.321  Mean   :3.554  Mean   :6.31
3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:5.00  3rd Qu.:4.000  3rd Qu.:4.000  3rd Qu.:2.000  3rd Qu.:3.000  3rd Qu.:5.000  3rd Qu.:5.000  3rd Qu.:8.00
Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :5.00  Max.   :5.000  Max.   :5.000  Max.   :5.000  Max.   :5.000  Max.   :5.000  Max.   :75.00
   G3        G1        G2
G3    Min. : 4.00  Min. : 3.00  Min. : 5.00
1st Qu.: 9.00  1st Qu.: 9.00  1st Qu.: 9.00
Median :11.00  Median :11.00  Median :11.00
Mean   :11.53  Mean   :11.27  Mean   :11.36
3rd Qu.:14.00  3rd Qu.:14.00  3rd Qu.:13.25
Max.   :20.00  Max.   :19.00  Max.   :19.00
```

Appendix 12: Relevant graphs

Variable: Romantic and G3

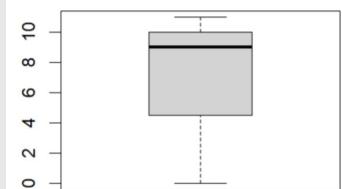


Variable: Studytime and G3

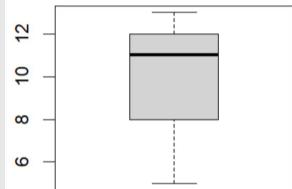


Variable: Social Factors and G3

Mean: 6.67



Mean: 9.67



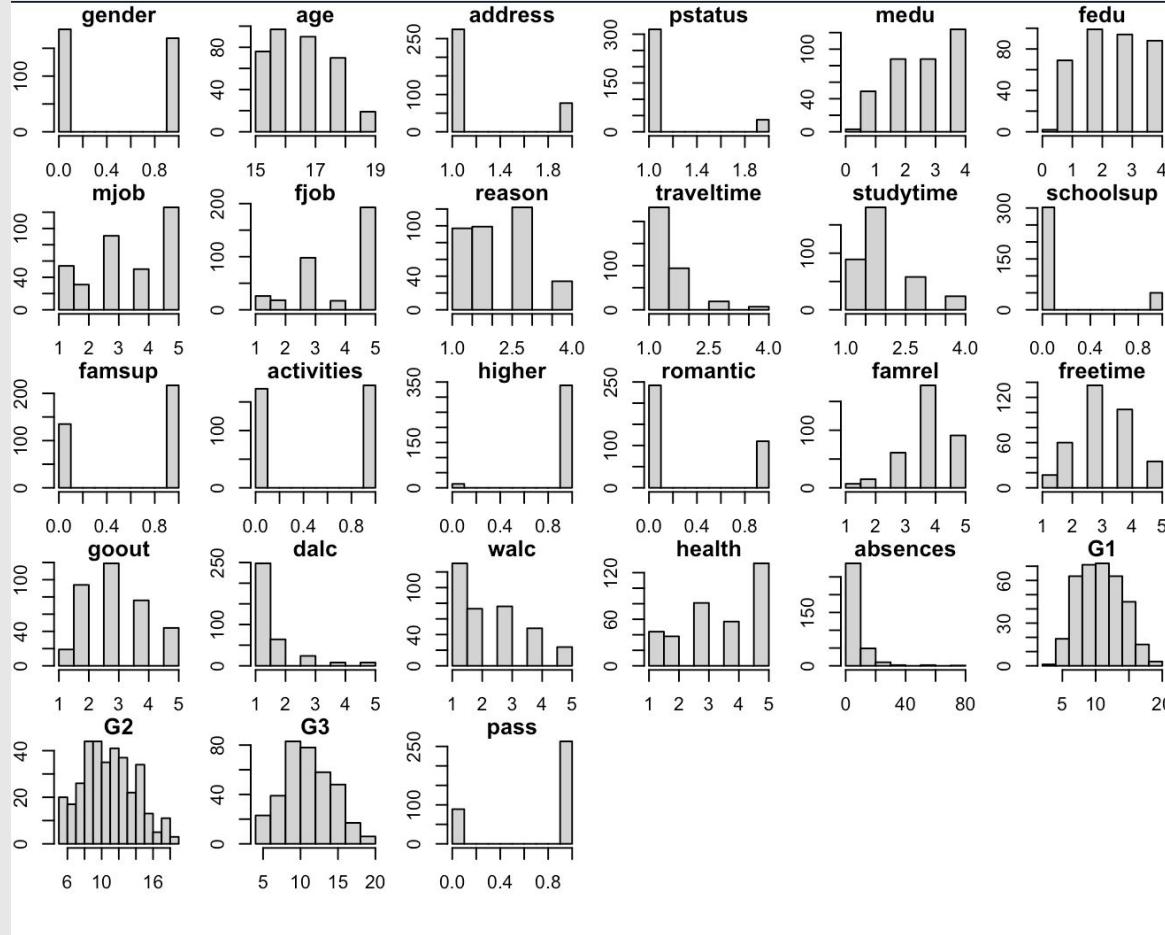
Variable: Correlation Matrix
School/Personal Factors and Higher

	address	higher	Medu	Fedu	Mjob	Fjob	G3	schoolsup	famsup
address	1	-0.05	-0.13	-0.07	0.09	0.05	-0.1	-0.03	-0.02
higher	-0.05	1	0.18	0.17	-0.08	0.01	0.18	0.05	0.09
Medu	-0.13	0.18	1	0.63	-0.53	-0.22	0.21	-0.04	0.18
Fedu	-0.07	0.17	0.63	1	-0.34	-0.37	0.15	0.03	0.18
Mjob	0.09	-0.08	-0.53	-0.34	1	0.31	-0.14	0.1	-0.14
Fjob	0.05	0.01	-0.22	-0.37	0.31	1	-0.1	-0.07	-0.05
G3	-0.1	0.18	0.21	0.15	-0.14	-0.1	1	-0.08	-0.04
schoolsup	-0.03	0.05	-0.04	0.03	0.1	-0.07	-0.08	1	0.1
famsup	-0.02	0.09	0.18	0.18	-0.14	-0.05	-0.04	0.1	1

Appendix 13: Confusion matrix for 70/30 split

Confusion Matrix (Testing Data) “1” → Pass “0” → Fail		
Prediction	Fail	Pass
Actual (Fail)	21	2
Actual (Pass)	2	81
Accuracy: 95% Recall: 89% Precision: 94% F1: 0.91		

Appendix 14: Distribution graphs checking for normality



Appendix 15: Data Link

```
32 x 1 sparse Matrix of class "dgCMatrix"
      s1
(Intercept) 6.65781710
(Intercept) .
schoolMS .
Gender     -0.53273224
age        -0.27316514
address    .
famsize   .
Pstatus   .
Medu       0.46786978
Fedu      .
Mjob      .
Fjob      .
traveltime -0.05237420
studytime  0.34519182
failures   -0.52804918
schoolsup  .
famsup    .
paid       0.63302182
activities .
nursery   .
internet  .
romantic  .
famrel    .
freetime   .
goout     .
Dalc      .
Walc      .
health    .
absences   -0.02275892
G1        .
G2        .
G3        .
```

Appendix 16: Data Link

<https://archive.ics.uci.edu/ml/datasets/student+performance#>