

# Advanced Statistics Assignment

Bruno Lenderink, Fabiana Caccavale, Marco Amadori, Lisa Aita

November 10, 2021

## temporary description of project

Import then into R the attached data.

Provide descriptives of relevant variables and associations in the data. Select then G3 as your target variable, and provide a linear regression model to predict it. Do not use G1 or G2 as potential predictors. Interpret the results of your final model, and assess its predictive performance. Your deliverables are a 2-3 page report (not counting tables and figures), and a replication package in the form of data and a text file with R code that I can copy and paste to reproduce your analysis.

#To remove all the objects that are stored in your global environment (but not remove any of loaded packages):

#Packages #To put all packages (that were previously installed) into memory:

Table 1: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
age	649	16.744	1.218	15	16	18	22
Medu	649	2.515	1.135	0	2	4	4
Fedu	649	2.307	1.100	0	1	3	4
traveltime	649	1.569	0.749	1	1	2	4
studytime	649	1.931	0.830	1	1	2	4
failures	649	0.222	0.593	0	0	0	3
famrel	649	3.931	0.956	1	4	5	5
freetime	649	3.180	1.051	1	3	4	5
goout	649	3.185	1.176	1	2	4	5
Dalc	649	1.502	0.925	1	1	2	5
Walc	649	2.280	1.284	1	1	3	5
health	649	3.536	1.446	1	2	5	5
absences	649	3.659	4.641	0	0	6	32
G1	649	11.399	2.745	0	10	13	19
G2	649	11.570	2.914	0	10	13	19
G3	649	11.906	3.231	0	10	14	19

[1] 5.865375

Table 2: Regression results using all variables

	<i>Dependent variable:</i>
	G3
schoolMS	-1.200*** (0.267)
sexM	-0.633** (0.250)
age	0.156 (0.102)
addressU	0.323 (0.262)
famsizeLE3	0.303 (0.245)
PstatusT	0.177 (0.347)
Medu	0.035 (0.151)
Fedu	0.167 (0.138)
Mjobhealth	0.901* (0.538)
Mjobother	0.050 (0.303)
Mjobservices	0.421 (0.373)
Mjobteacher	0.512 (0.502)
Fjobhealth	-0.612 (0.752)
Fjobother	-0.184 (0.456)
Fjobservices	-0.643 (0.479)
Fjobteacher	0.580 (0.672)
reasonhome	0.051 (0.285)
reasonother	-0.435 (0.368)
reasonreputation	0.218 (0.298)
guardianmother	-0.338 (0.265)
guardianother	0.105 (0.532)
traveltime	0.062 (0.159)
studytime	0.407*** (0.140)
failures	-1.412*** (0.205)
schoolsupyes	-1.311*** (0.364)
famsupyes	-0.020 (0.228)
paidyes	-0.372 (0.461)
activitiesyes	0.219 (0.223)
nurseryyes	-0.216 (0.271)
higheryes	1.733*** (0.383)
internetyes	0.253 (0.276)
romanticyes	-0.432* (0.229)
famrel	0.162 (0.116)
freetime	-0.138 (0.112)
goout	-0.066 (0.107)
Dalc	-0.205 (0.153)
Walc	-0.081 (0.118)
health	-0.187** (0.077)
absences	-0.038 (0.025)
Constant	8.681*** (1.985)
Observations	649
R <sup>2</sup>	0.360
Adjusted R <sup>2</sup>	0.319
Residual Std. Error	2.665 (df = 609)
F Statistic	8.797*** (df = 39; 609)

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

Table 3: Regression results using AIC

	<i>Dependent variable:</i>
	G3
schoolMS	−1.513*** (0.240)
sexM	−0.571** (0.236)
age	0.167* (0.099)
Medu	0.301*** (0.099)
guardianmother	−0.453* (0.253)
guardianother	0.034 (0.512)
studytime	0.409*** (0.135)
failures	−1.484*** (0.198)
schoolsupyes	−1.336*** (0.357)
higheryes	1.864*** (0.377)
romanticyes	−0.422* (0.225)
Dalc	−0.358*** (0.123)
health	−0.180** (0.074)
absences	−0.037 (0.024)
Constant	8.905*** (1.757)
Observations	649
R <sup>2</sup>	0.334
Adjusted R <sup>2</sup>	0.319
Residual Std. Error	2.666 (df = 634)
F Statistic	22.705*** (df = 14; 634)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 4: Regression model using AIC selected variables on only training data

	<i>Dependent variable:</i>
	G3
schoolMS	−1.635*** (0.278)
sexM	−0.637** (0.272)
age	0.246** (0.115)
Medu	0.318*** (0.114)
guardianmother	−0.499* (0.291)
guardianother	0.008 (0.571)
studytime	0.462*** (0.152)
failures	−1.421*** (0.228)
schoolsupyes	−1.299*** (0.410)
higheryes	1.791*** (0.436)
romanticyes	−0.444* (0.258)
Dalc	−0.252* (0.141)
health	−0.123 (0.084)
absences	−0.043 (0.027)
Constant	7.228*** (2.033)
Observations	519
R <sup>2</sup>	0.323
Adjusted R <sup>2</sup>	0.305
Residual Std. Error	2.733 (df = 504)
F Statistic	17.209*** (df = 14; 504)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01