NOVA IMS

Information Management School

Master Program in Information Management

Understanding Students' Academic Achievement

Evidence for Portugal

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa

Index

Understanding Students' Academic Achievement

Introduction

timeless topic human capital new businesses and technologies

information and knowledge variation of salaries

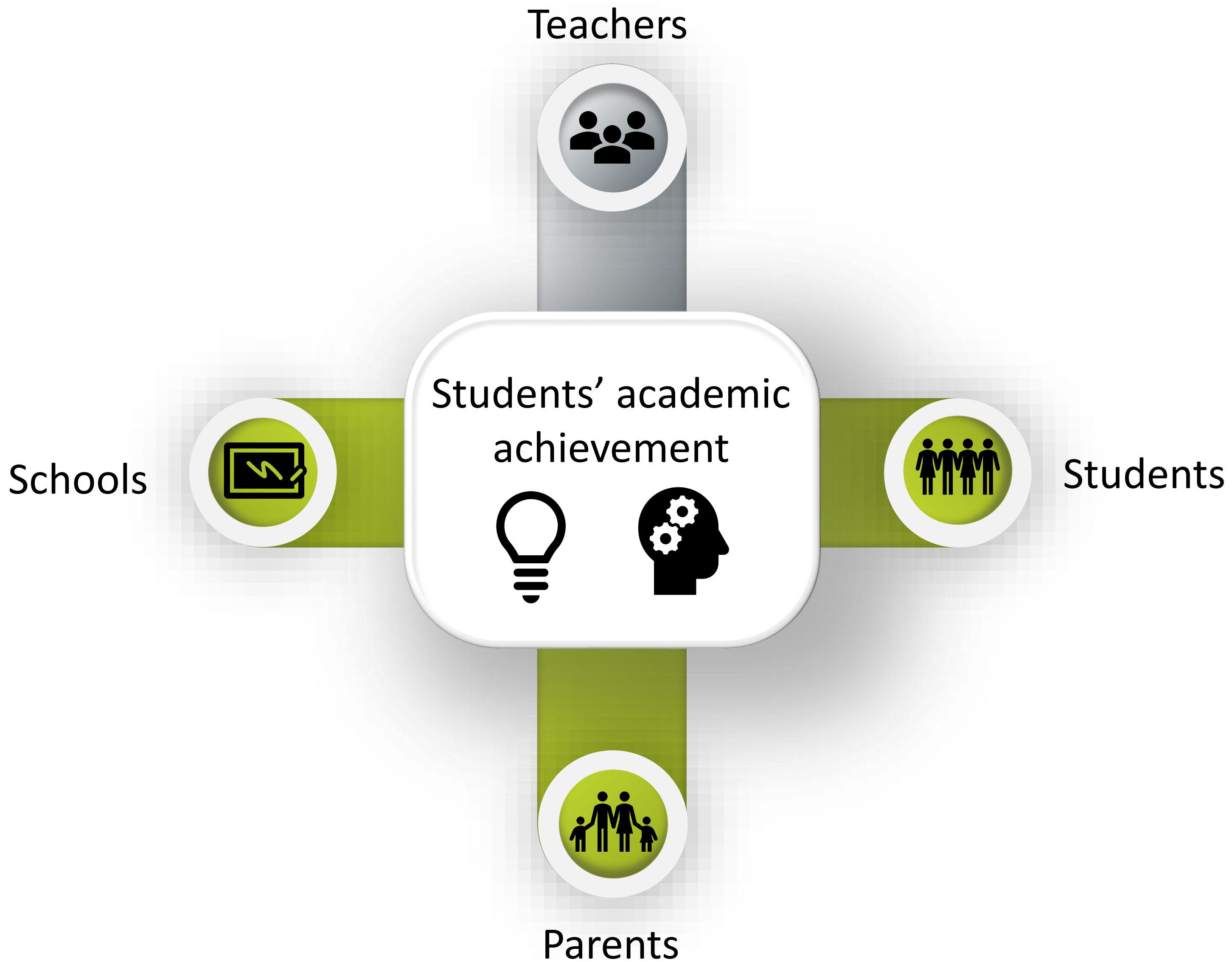
country development social exclusion

discrimination of minority groups education world of work

Research Question

Which factors affect students' academic achievement ?

Theoretical Background



Theoretical Background - Research

Ref	Methods	Students	Parents	Schools	Teachers
(Hanushek & Kimko, 2000)	Regression models	x		x	
(Hoxby, 2000)	Regression models	x		x	
(Fan & Chen, 2001)	General linear model (GLM)	x	x		
(Barnett et al., 2002)	Linear Programming techniques			x	
(Rockoff, 2004)	Regression models				x
(Driessen et al., 2005)	Frequency, Variance and Structural models	x	x	x	
(Rivkin et al., 2005)	Regressions models			x	x
(Archibald, 2006)	Hierarchical linear models (HLM)	x		x	x
(Jackson et al., 2006)	Internet recorded	x			
(J.-S. Lee & Bowen, 2006)	Hierarchical linear models (HLM)	x	x		
(Marks et al., 2006)	Item Response Theory (IRT) Regression models	x	x	x	
(Jeynes, 2007)	Regression models		x		
(Codjoe, 2007)	Interviews	x			
(Croninger et al., 2007)	Hierarchical linear models (HLM)	x			x
(H. Lee, 2007)	Hierarchical linear models (HLM) Classic lineal regression model	x	x	x	
(Lei & Zhao, 2007)	Hierarchical linear models (HLM) ANOVA tests	x			
(Steinmayr & Spinath, 2008)	Regressions models	x			
(Caro et al., 2009)	Hierarchical linear models (HLM) Panel data models	x			

Theoretical Background - Research

(Mensah & Kiernan, 2010)	Tobit regression models Univariate and Multivariate analyses	x	x		
(Hanushek, 2011)	Regression models				x
(Hartas, 2011)	Univariate analyses of variance Chi-square tests		x		
(Patterson & Pahlke, 2011)	Regression models	x	x		
(Hanushek & Woessmann, 2012)	Regression models	x		x	
(Brunner et al., 2013)	Multiple group factor analytic models Full maximum likelihood method "MLR"	x			
(Wally-Dima & Mbekomize, 2013)	Descriptive statistics T tests	x			
(Bosworth, 2014)	Regression models	x		x	
(Krassel & Heinesen, 2014)	Regression discontinuity design (RDD) Control for school fixed effects (SFE) Ordinary Least Squares (OLS)	x	x	x	
(Vigdor et al., 2014)	Probit regression Regression models	x			
(Hodis et al., 2015)	Hierarchical linear models (HLM)	x			
(C. L. Lee & Mallik, 2015)	Ordinary Least Squares (OLS)	x			

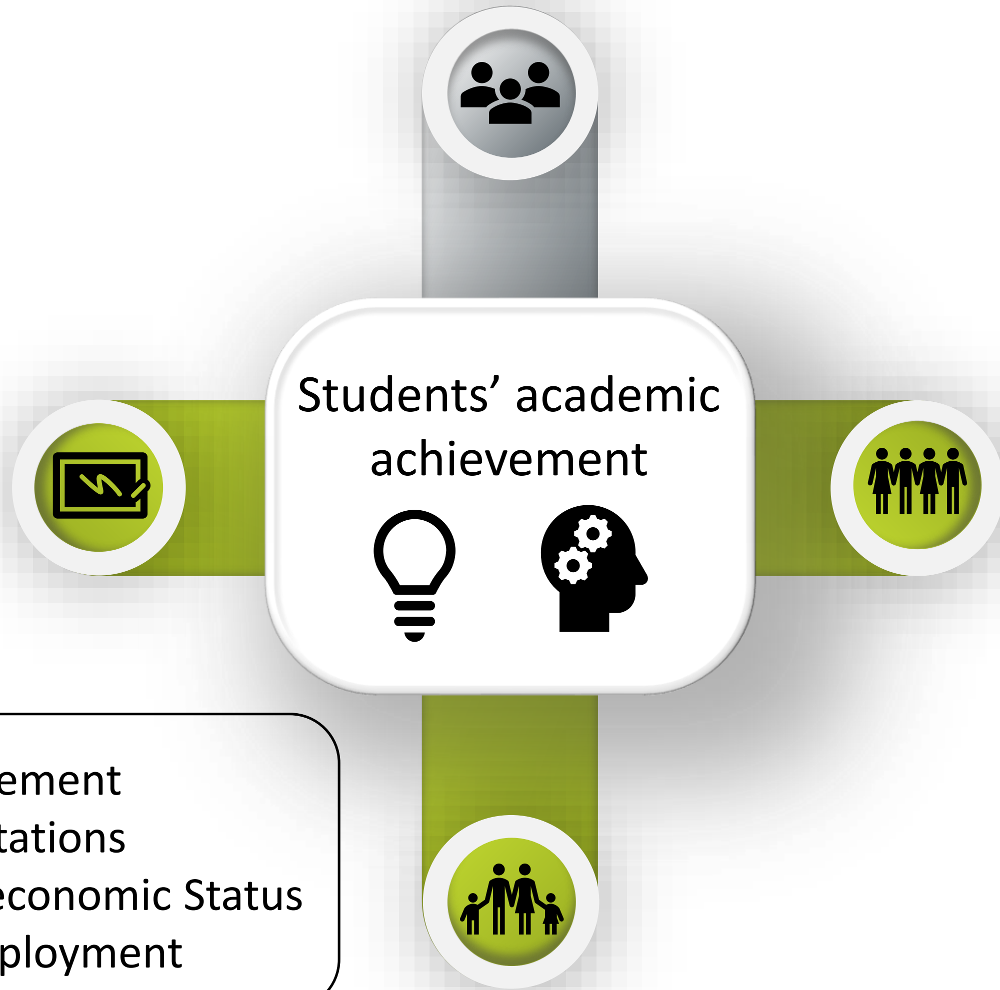
Students	24 papers
Parents	10 papers
Schools	11 papers
Teachers	5 papers

Theoretical Background

Students

- Gender
- Ethnicity
- Computer access
- Internet access

Theoretical Background



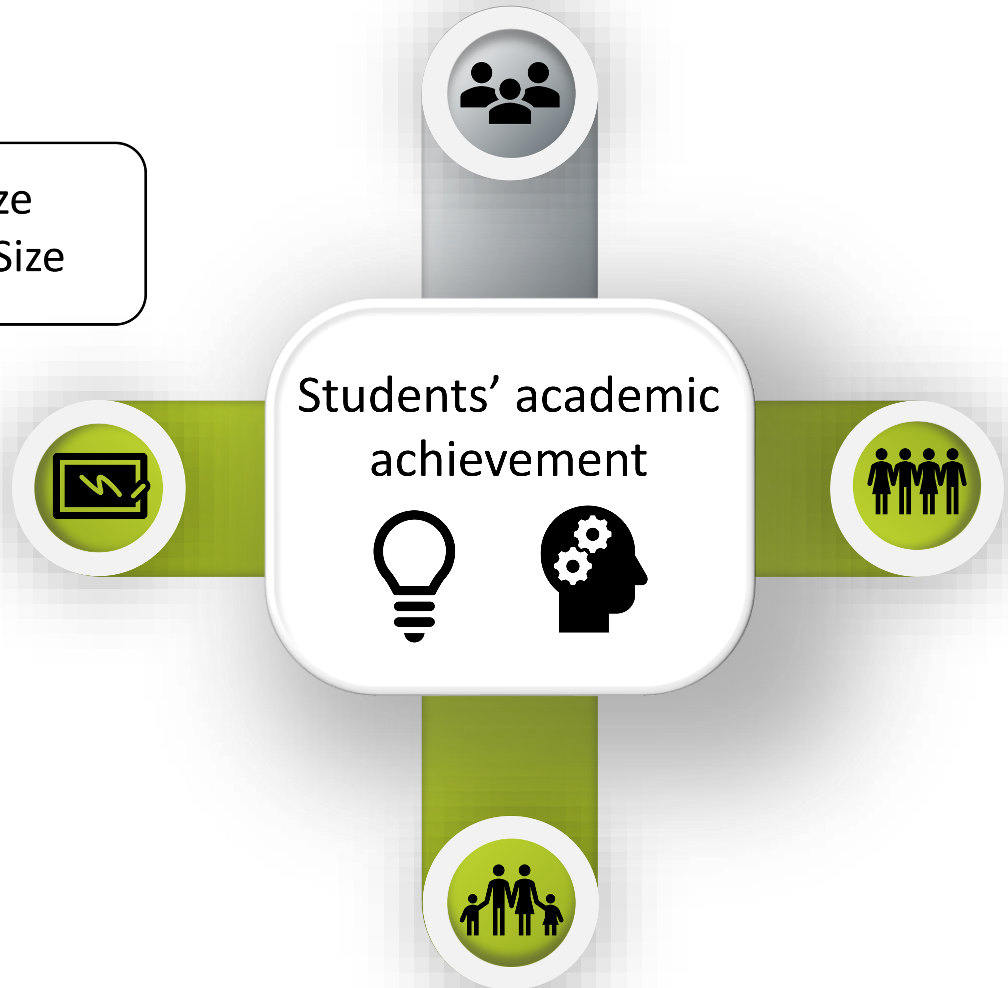
- Involvement
- Expectations
- Socioeconomic Status
- Unemployment

Parents

Theoretical Background

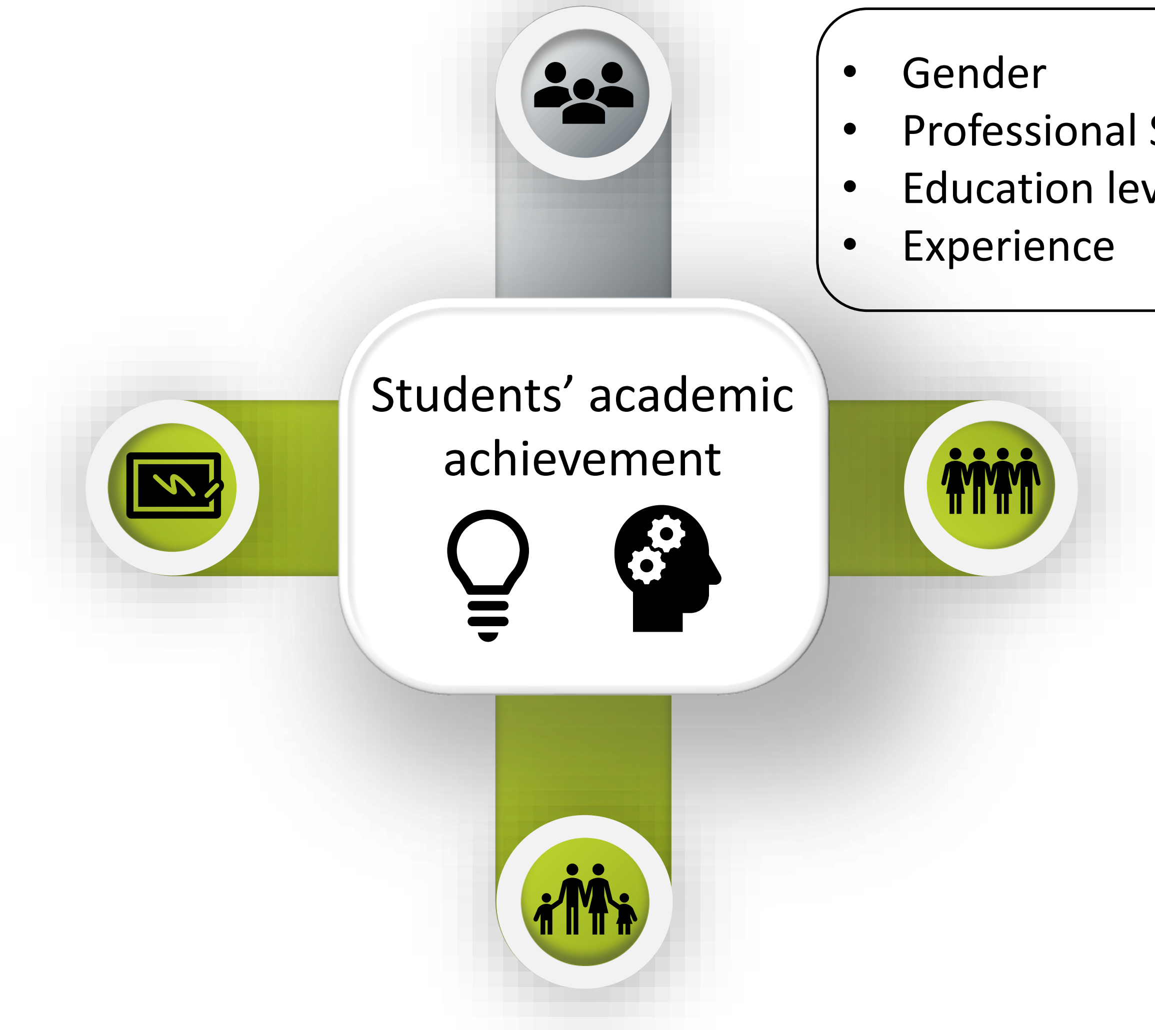
- Class Size
- School Size

Schools



Theoretical Background

Teachers



- Gender
- Professional Situation
- Education level
- Experience

Conceptual Model

Students' Variables:

H1

- Gender will have an impact on students' academic achievement as females will perform better

H2

- Native students will perform better on academic achievement

H3

- Students with computer access will perform better on academic achievement

H4

- Students with internet access will perform better on academic achievement

H5

- Students that have reproved in the past will present lower levels on academic achievement in the future

Parents' Variables:

H6

- Students that receive support from social services (SASE) will have lower levels on academic achievement

H7

- Students that receive family financial support will have lower levels on academic achievement

H8

- Parents education level will have a positive impact on academic achievement

Schools' Variables:

H9

- Class size will have a negative impact on academic achievement

H10

- School size will have a positive impact on academic achievement

Methodology - Data

November 2016

January 2017

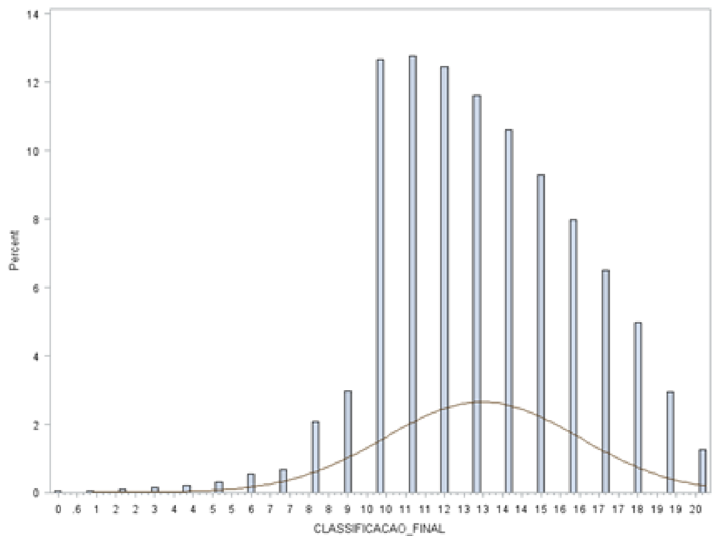
May 2017

- DGEEC – MISI Database
  - Students from: Portuguese public schools in 2014/2015 from 10th, 11th and 12th grades evaluated and attending the 21 courses
- INE
  - Population density
  - Monthly average income
  - Percentage average on culture expenses
  - Aging index
  - Residence population
  - Unemployment rate

383560 observations

Methodology – Non-Parametric tests

Dependent variable – Final Classification doesn’t follow a normal distribution



Test for Normality		
Test	Statistic	P-Value
Kolmogorov-Smirnov	D = 0,099641	Pr<D <0,0100

Results – Non-Parametric tests

Students characteristics

Variables	n	%	Mean	SD	Mann-Whitney /Kruskal-Wallis (k)	Conover Variance Test
Gender						
Female	62.174	55.9%	13.38	2.98	-2548.4738***[1]	-2.5849***
Male	49.128	44.1%	12.90	3.04		
Age (k)						
[0-16[	153	0.2%	13.98	3.46	14072.9362***	3777.8620***
[16-18[	90.682	81.5%	13.36	2.99		
[19-21[	19.731	17.7%	11.45	2.67		
]>=21]	736	0.7%	11.74	3.66		
N_Reprov by year						
10th,0 rep	10.475	9.4%	13.18	3.15	4050.3054***	372.5845***
10th,1 rep	2.870	2.6%	11.40	2.93		
10th,2 reps	187	0.2%	10.82	2.91		
10th, +2reps	209	0.2%	12.08	2.95		
11th,0 rep	36.124	32.5%	13.77	2.89	16579.2848***	2644.0207***
11th,1rep	13.116	11.8%	11.95	2.65		
11th,2reps	976	0.9%	11.09	2.64		
11th, +2reps	444	0.4%	11.08	3.72		
12th,0 rep	31.725	28.5%	13.64	2.88	9025.3129***	1871.5462***
12th,1 rep	13.990	12.6%	11.61	2.67		
12th,2 reps	785	0.7%	10.93	2.86		
12th, +2 reps	401	0.4%	11.83	3.51		
Nationality						
Portugal	108.134	97.2%	13.19	3.01	730.1785***	-6.0427***
Other	3.168	2.8%	12.37	3.01		

For a significance level of 1%, we reject the null hypothesis (p-value <0.0001)

Results – Non-Parametric tests

Parents' Socioeconomics characteristics

Variables	n	%	Mean	SD	Kruskal-Wallis	Conover Variance Test
Beneficiary_SASE						
No Support	81.787	73.5%	13.33	3.05		
Level 1 (Highest support)	15.215	13.7%	12.89	2.88	3119.6251***	1080.8923***
Level 2 (Highest support)	14.300	12.8%	12.59	2.87		
Family Financial support (FFS)						
No Support	81.406	73.1%	13.32	3.05		
Level 1 (Highest support)	13.351	12.0%	12.59	2.85	2920.3323***	1151.4959***
Level 2 (Medium support)	15.242	13.7%	12.89	2.88		
Level 3 (Lowest support)	1.303	12.0%	13.22	2.94		

For a significance level of 1%, we reject the null hypothesis (p-value <0.0001)

Results – Non-Parametric tests

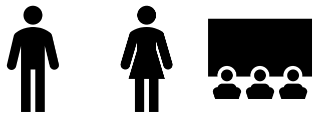
Schools characteristics

Variables	n	%	Mean	SD	Kruskal-Wallis	Conover Variance Test
Class Size						
[1-10]	864	0.8%	12.66	2.93		
]10-20]	14.389	12.9%	13.01	2.93		
]20-25]	25.974	23.3%	13.16	2.97	971.5192***	588.0421***
]25-30]	52.046	46.8%	13.29	3.05		
]30-40]	17.152	15.4%	12.93	3.02		
]>=40[	877	0.8%	12.15	3.44		
School Size						
[1-100]	2.469	2.2%	12.83	3.14		
]100-200]	8.954	8.0%	12.85	3.01		
]200-300]	9.293	8.3%	12.93	3.02		
]300-500]	24.223	21.8%	13.11	2.92		
]500-600]	10.073	9.1%	13.34	3.01	1092.5769***	634.9647***
]600-700]	14.119	12.7%	13.30	3.01		
]700-900]	11.762	17.8%	13.12	3.08		
]>=900[	22.352	20.1%	13.34	3.04		

For a significance level of 1%, we reject the null hypothesis (p-value <0.0001)

Results – Non-Parametric tests

Courses by Gender



Variables	n	%	Mean	SD	Kruskal-Wallis	Conover Variance Test
Courses by Gender						
Drawing A, Female	4.359	1.1%	14.75	2.32	89.1649***	1.9836***
Drawing A, Male	2.021	0.5%	14.14	2.42		
Philosophy, Female	32.715	8.5%	13.76	2.84	1038.7988***	-6.2612***
Philosophy, Male	25.705	6.7%	13.00	2.83		
History A, Female	13.506	3.5%	12.50	2.66	104.3030***	-8.9812***
History A, Male	6.465	1.7%	12.09	2.54		
Foreign Language – English, Female	32.701	8.5%	14.41	3.16	15.4312***	-14.2941***
Foreign Language – English, Male	26.645	7.0%	14.51	3.00		
Mathematics A, Female	23.757	6.2%	13.01	3.49	314.4330***	-1.5765***
Mathematics A, Male	23.302	6.1%	12.43	3.56		
Portuguese, Female	43.285	11.3%	13.20	2.47	2508.7112***	-5.6441***
Portuguese, Male	33.148	8.6%	12.30	2.46		

For a significance level of 1%, we reject the null hypothesis (p-value <0.0001)

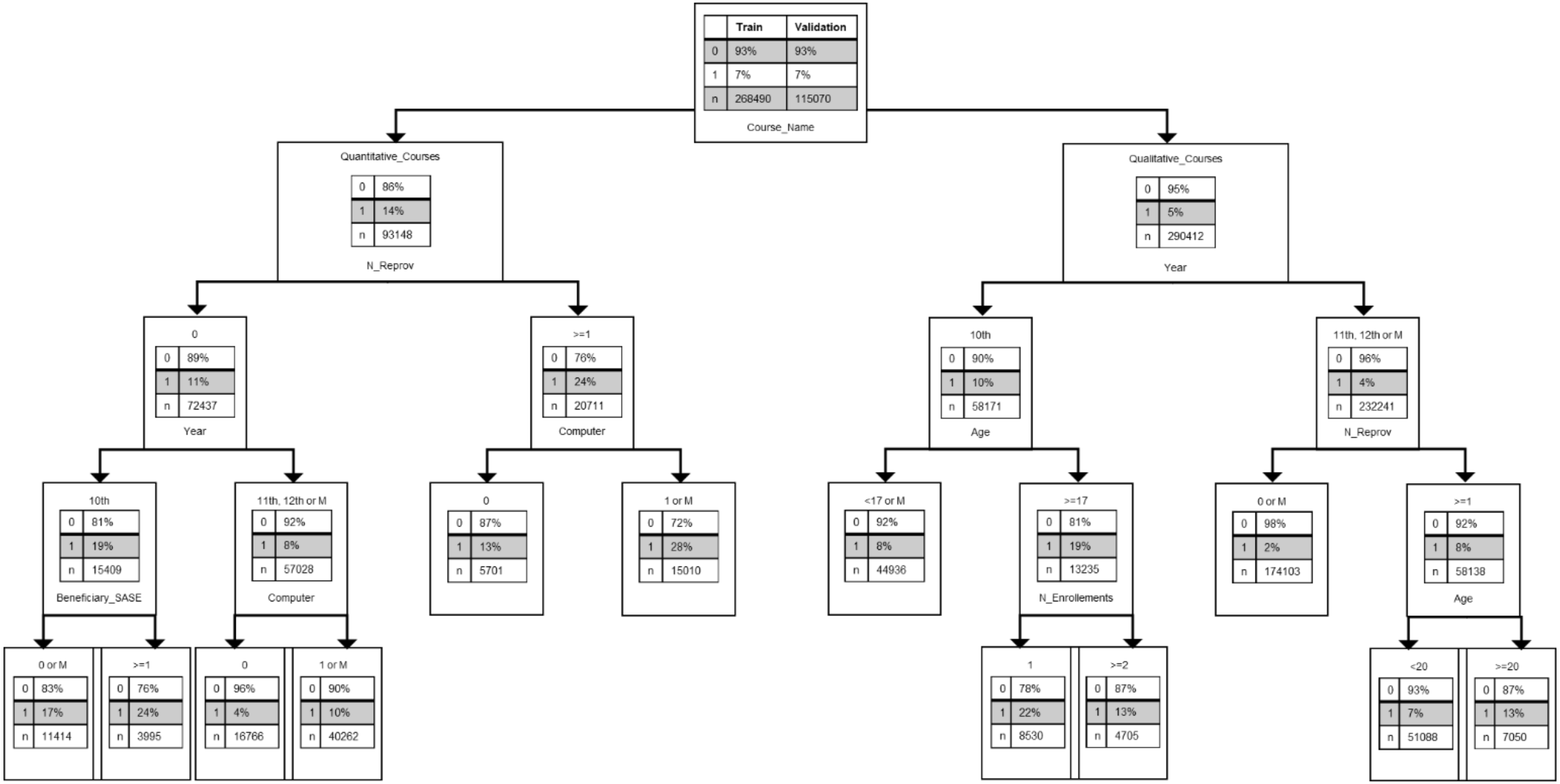
Methodology – Decision Trees

Data Mining – Decision Trees

Model 1 - Decision Tree for academic achievement at course level
Target variable: Student passed or not per course

Model 2 - Decision Tree for academic achievement at year level
Target variable: Final grade >=10 or <10 per academic year

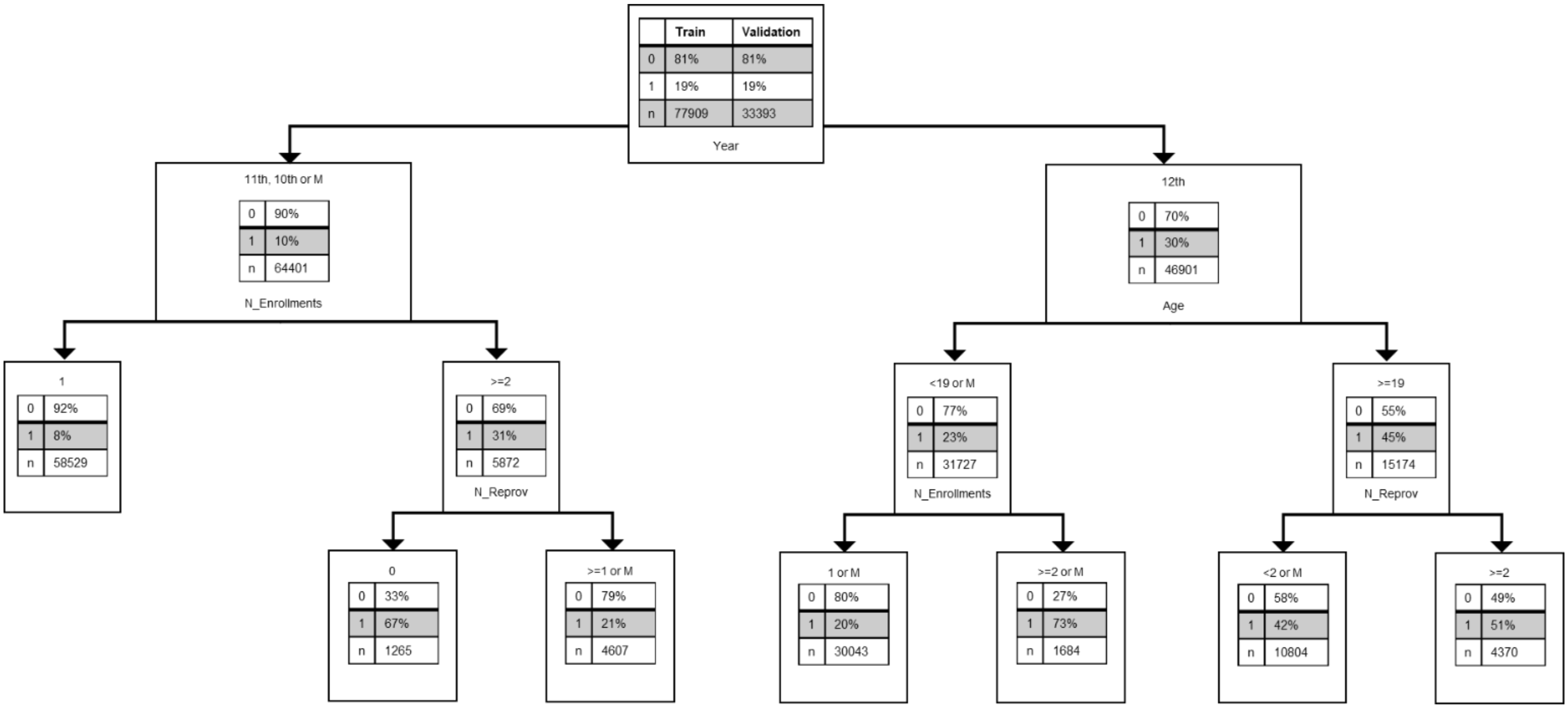
Results - Decision Tree - Model 1



- Quantitative Courses: Descriptive Geometry A, Mathematics A, Mathematics applied to Social Sciences, Mathematics B, Physics and Chemistry A, Portuguese as non Maternal Language
- Qualitative Courses: Biology and Geology, Drawing A, Economy A, Foreign Language – English, Foreign Language – French, Foreign Language – German, Foreign Language – Spanish, Geography A, History A, History B, History of Culture and Arts, Latin A, Philosophy, Portuguese, Portuguese Literature

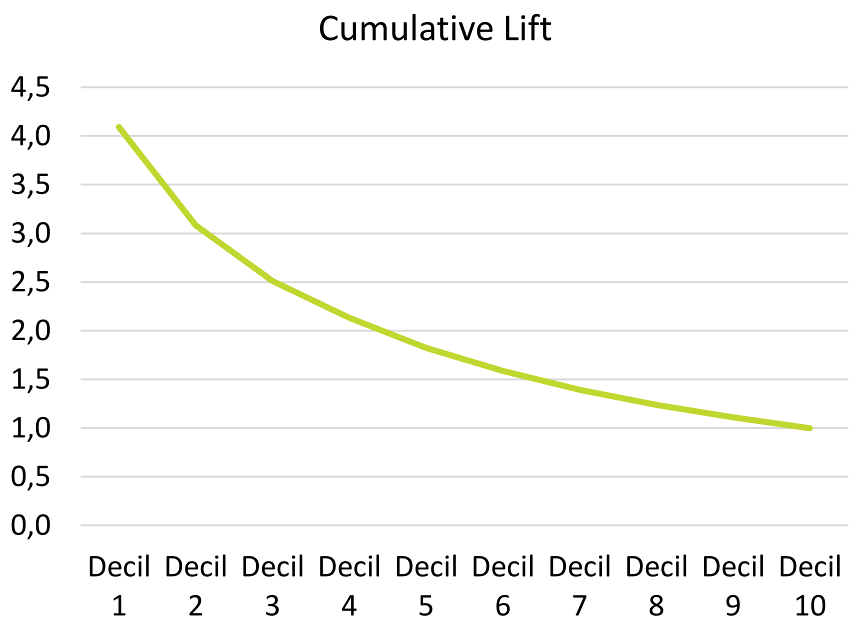
0 – Students with positive approval rate; 1 – Students with reprove rate

Results - Decision Tree - Model 2

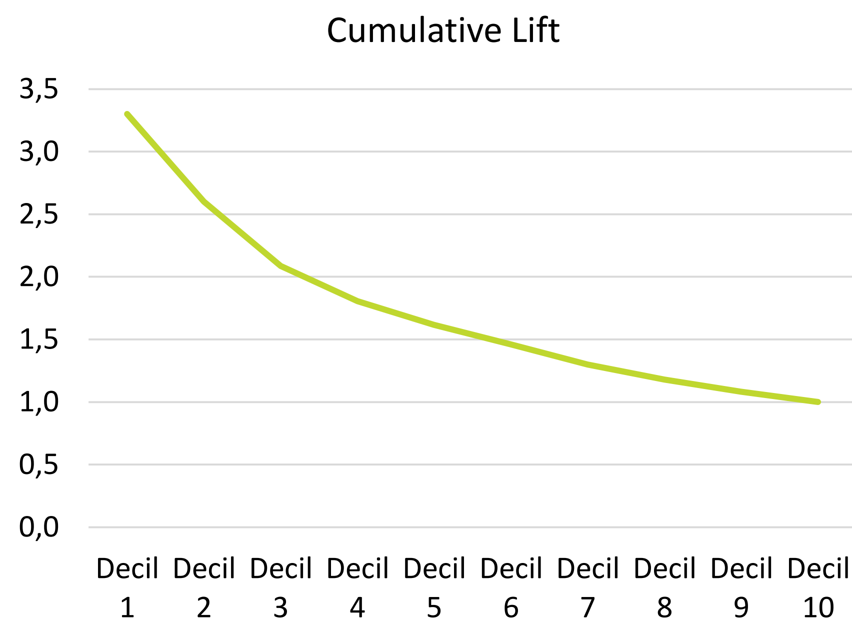


0 – Students with positive approval rate; 1 – Students with reprove rate

Results – Cumulative Lift



Cumulative Lift for Model 1



Cumulative Lift for Model 2

Discussion - Findings

01

Academic Achievement at course Level

- Course

Academic Achievement at year Level

- Year enrolled (10th, 11th, 12th)

02

03

All the hypothesis studied are statistically significant to students' academic achievement, except H8 where we could not infer any conclusion.

Practical and Theoretical Implications

Practical

First
Investing on a society and education with better digital and technological networks

Second
Help the group of students with less income

Third
Invest on the reduction of the number of students per class

Fourth
Know students background to what refers to the number of reproves or good performance

Theoretical

First
Data mining techniques proved to yield good results

Second
Non-parametric tests appear to be a good alternative to parametric tests

Conclusions

Students reprove background still has a great emotional weight on academic achievement

Access to computers and Internet, when well used for school purposes, are a powerful mean

Students coming from less wealthy households obtain lower scholar performances

There are still differences on academic achievement among female and male genders

NOVA IMS Information Management School

Limitations and Future Work

Limitation		Future work
Data Quality: Data pre-processing took considerable time; Missing values and data inconsistency are aspects to improve	→	Further developments are needed in the way data is recorded and stored
Data used is cross-sectional	→	It would be interesting to do an analysis on academic achievement for multiple points in time
Recurred to secondary data	→	Include other potential antecedents/variables of academic achievement and consequently other type of methods

Thank you!

By: Ana Filipa Rosa Louro

Advisor: Professor Doutor Frederico Miguel de Campos Cruz Ribeiro de Jesus
Co Advisor: Engenheiro Jorge Nelson Gouveia de Sousa Neves

Address: Campus de Campolide, 1070-312 Lisboa, Portugal
Phone: +351 213 828 610 Fax: +351 213 828 611

Acreditações e Certificações

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa