An analysis of the differences in Italian students' performance in STEM and no-STEM courses

Michele La Rocca, Marcella Niglio and Marialuisa Restaino

Dept of Economics and Statistics (DiSES) & Laboratory of Statistics (StatLab)
University of Salerno

MESE1
Measurement in STEM Education 1.
Napoli, 30 January - 1 February 2023

Structure of the talk

1. Motivation and Research questions
2. Data on students' performance
3. Model strategy
4. Results of data analysis
   - Exploratory data analysis
   - Results of logistic regression
   - Results of ordinal logistic regression
5. Conclusions
6. References
7. Appendix
   - Logistic regression plots
   - Ordinal logistic regression plots

Structure of the talk

Motivation and Research questions

- The evaluation of educational services attracts scholars and lots of papers have been published over the years.
- The main educational services and phenomenon under investigation are
  - students'/staff/professors' mobility;
  - students' performance;
  - degree and time for graduation;
  - students' dropouts;
  - Student Evaluation of Teaching (SET);
  - quality of research.

Motivation and Research questions

- The evaluation of educational services attracts scholars and lots of papers have been published over the years.
- The main educational services and phenomenon under investigation are
  - students'/staff/professors' mobility;
  - students' performance;
  - degree and time for graduation;
  - students' dropouts;
  - Student Evaluation of Teaching (SET);
  - quality of research.

Motivation and Research questions

- Predicting students' academic performance is a key step in order to improve the efficiency of university systems.
- Universities rely on information about the high school career, e.g. type of school and various measures of proficiency.
- The results at high school are not fully appropriate to predict academic performance:
  a) mismatch between competencies evaluated at high school and competencies required for a given degree program;
  b) heterogeneity in the criteria for awarding marks (variability across types of schools and across geographical regions).
- Pre-enrolment assessment tests have been introduced, and they might facilitate faculties and departments to reduce the knowledge gap.
- The empirical research about predicting students' academic performance is scattered in various journals, ranging from Psychology to Economics.

Motivation and Research questions

- Recent literature offers several papers about student performance and its determinants, and results are not always in the same direction.
- A classification of papers could distinguish between
  - papers accounting for students' social and demographic characteristics (among others, [Boscaino et al., 2007], [Grilli et al., 2013], [Grilli et al., 2015]); and
  - papers accounting for their previous performance and/or psychological and subjective features (among others, [Adelfio et al., 2014], [Attanasio et al., 2013]).
- The literature generally suggests that the impact of the determinants varies (in terms of extent and direction) according to the context (economic, social, political, demographic, etc.) and results should hold just in that context.

Motivation and Research questions

- Within this framework, we analyze the differences in students' performance.
- In particular, we focus on students' performance, measured as the total number of gained credits (ECTS) after one year.
- In particular,
  - we focus on the threshold of 40 ECTS (Annual Review Report by ANVUR - Italian National Agency for the Evaluation of Universities and Research Institutes).
  - we figure out which background students' characteristics might influence their performance.
- The main research questions are:
  a) Does the pre-enrollment test score affect the probability of reaching the 40 ECTS during the first academic year?
  b) Is a gender gap detectable?
  c) Are there any differences among the courses of study?

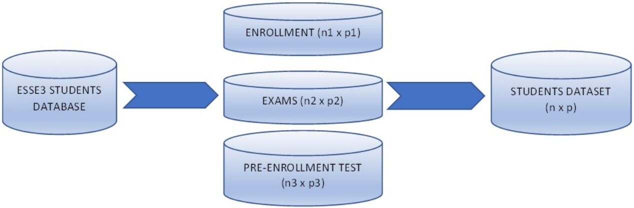
Structure of the talk

Data on students' performance

- The data refers to students who enroll at the University of Salerno.
- Among all departments, we collect information for Dept of Economics and Statistics (DiSES).
- We extract information from the Student Information System (ESSE3), which manages the entire career of students from enrollment to graduation.
- It contains information about students' high school diplomas, personal characteristics, exams, abroad experience, internship, and degrees.

Data on students' performance

- For the talk purposes, from this source, we collect and merge information on students' enrollment, exams, and pre-enrollment tests.
- The keys used for merging these datasets are the students' ID (*matricola*) and the fiscal code.



Data on students' performance

- The datasets considered are:
  1) "*Student enrolled*" (Iscritti), which contains a record for each student, with all background information (date of birth, place of residence, high school diploma, diploma mark, and so on).

$$\underset{(n_1 \times p_1)}{\mathrm{X}_{\textsf{student}}}.$$

  2) "*Student exams*" (Esami sost/non sost), which contains more records for a student since each record is related to each exam that should be taken by the student in the academic year (date of exam, exam mark, ECTS, and so on).

$$\underset{(n_2 \times p_2)}{\mathrm{X}_{\textsf{exam}}}.$$

  3) *pre-enrollment test*, which contains the test score of all students who took it. It consists of 36 questions (13 questions on Logic; 13 questions on Reading; 10 questions on Mathematics).

$$\underset{(n_3 \times p_3)}{\mathrm{X}_{\textsf{test}}}.$$

- The datasets' sizes are different. In particular:

$$n_2 >> n_1 \qquad \text{and} \qquad n_3 > n_1.$$

Data on students' performance

- A merging process consists of 4 major stages.
  1. transforming the "student exams" dataset into a new dataset, where there would be a record for each student. The main variables considered are
     - the total number of credits (ECTS) earned during the first year;
     - the total number of exams passed during the first year;
     - the mean of exam marks.

     $$\underset{(n_1 \times p_2^*)}{\mathrm{X}_{\text{exam.student}}} \subseteq \underset{(n_2 \times p_2)}{\mathrm{X}_{\text{exam}}}$$

     where $p_2^* < p_2$.
  2. merging the two datasets $\mathrm{X}_{\text{student}}$ and $\mathrm{X}_{\text{exam.student}}$, by the students' ID (*matricola*).

     $$\underset{(n_1 \times (p_1 + p_2^*))}{\mathrm{X}^*} = \underset{(n_1 \times p_1)}{\mathrm{X}_{\text{student}}} \cup \underset{(n_1 \times p_2^*)}{\mathrm{X}_{\text{exam.student}}}$$

  3. selecting the students who took the test and enrolled in the courses of study under investigation

     $$\underset{(n_1 \times p_3)}{\mathrm{X}_{\text{test.student}}} \subseteq \underset{(n_3 \times p_3)}{\mathrm{X}_{\text{test}}}$$

  4. merging X^* and $\mathrm{X}_{\text{test.student}}$, by the students' fiscal code.

     $$\underset{(n_1 \times p)}{\mathrm{X}} = \underset{(n_1 \times (p_1 + p_2^*))}{\mathrm{X}^*} \cup \underset{(n_1 \times p_3)}{\mathrm{X}_{\text{test.student}}}$$

where $p = (p_1 + p_2^* + p_3)$.

Data on students' performance

- The analysis covers four academic years (2018-2019 to 2021-2022).
- We consider the students who have enrolled in one of the first-level courses of study offered by the Department of Economics & Statistics (University of Salerno):
  - Business Administration (BA - L18 - code 02121);
  - Economics (E - L33 - code 02124);
  - Administration and Organization (A&O - L16 - 12122);
  - Tourism Management and Development (TM&D - L15 - 02125);
  - Statistics for Big Data (SBD - L41 - 02128).

Data on students' performance

- Given the heterogeneity of pre-enrollment tests, we focus on the courses for which the TOLC-E is required:
  - Business Administration (BA - L18 - code 02121);
  - Economics (E - L33 - code 02124);
  - Administration and Organization (A&O - L16 - 12122);
  - Tourism Management and Development (TM&D - L15 - 02125);
  - Statistics for Big Data (SBD - L41 - 02128).
- Among all variables available in ESSE3 and TOLC-E, we have considered the variables described in the next Table.

Data on students' performance

Variable	Short Description	Type
Academic Year	Academic Year of enrollment	Nominal
Course of study	Course of study where students have enrolled (BA, E, SBD)	Nominal
Gender	Gender of the graduated student	Nominal
Type of diploma	High school type (Classical studies, Technical, Scientific, ...)	Nominal
Diploma mark	Total marks of high school diploma (from 60 to 100)	Integer
Logic score	Pre-enrollment test - Logic	Integer
Reading score	Pre-enrollment test - Reading	Integer
Mathematics score	Pre-enrollment test - Mathematics	Integer
Total Test score	Pre-enrollment test - Total mark	Integer
AER (OFA)	Additional Educational Requirements (1=Yes vs 0=No)	Nominal

Structure of the talk

Model strategy

- The interest is in studying
  1. a binary variable $\Longrightarrow$ Logistic regression

$$Y = \begin{cases} 1 & \text{ECTS} \geq 40 \\ 0 & \text{otherwise} \end{cases} \Longrightarrow \Pr(Y = 1|\mathbf{x}) = \frac{\exp\left(\beta_0 + \mathbf{x}'\boldsymbol{\beta}\right)}{1 + \exp\left(\beta_0 + \mathbf{x}'\boldsymbol{\beta}\right)}$$

  2. an ordinal variable $\Longrightarrow$ Ordinal Logistic regression

$$Y = \begin{cases} 1 & \text{ECTS} = 0 \\ 2 & 0 < \text{ECTS} \leq 39 \\ 3 & 40 \leq \text{ECTS} < 60 \\ 4 & \text{ECTS} = 60 \end{cases} \Longrightarrow \Pr(Y \geq j|\mathbf{x}) = \frac{\exp\left(\beta_{0j} + \mathbf{x}'\boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{0j} + \mathbf{x}'\boldsymbol{\beta}\right)}$$

  for $j = 1, 2, 3, 4.$
- The aim is to identify which student's characteristics might affect the probability of reaching the threshold of 40 ECTS.

$\Longrightarrow$ We only consider the first year of study, since the threshold of 40 ECTS is relevant for some indicators which evaluate the efficiency of courses of study during each first year of study.

Model strategy

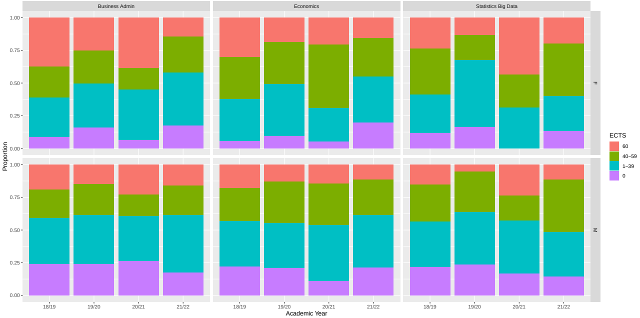
- We recall the main research questions:
  a) Does the pre-enrollment test score affect the probability of reaching the 40 ECTS during the first academic year?
  b) Is a gender gap detectable?
  c) Are there any differences among the courses of study?

Structure of the talk

Structure of the talk

Exploratory data analysis

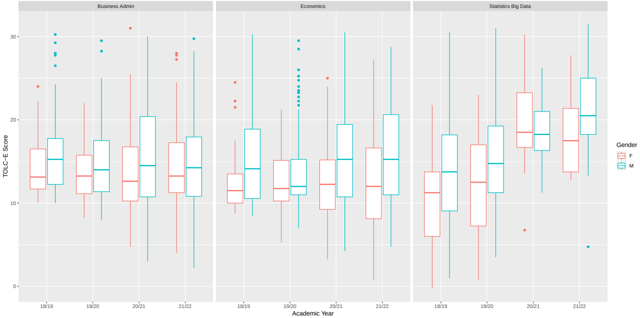
Figure: Barplot of ECTS for Courses of study, Academic years, and Gender



- The proportion of females who earn more than 40 ECTS is greater than that of males.
- Males have a higher proportion for the two lower classes.
- Students in Business Administration seem to have better performance than those in Economics.
- Students in Statistics improve their performance after the introduction of selective tests in 2020-2021.

Exploratory data analysis

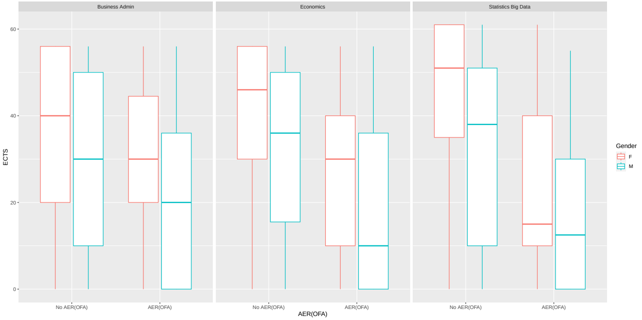
Figure: Boxplot of Tolc-E for Courses of study, Academic years, and Gender



- Men have a higher score.
- Economics shows a lower score than Business Administration in 2018-19 and 2019-20, while in the other two years, it has a better performance.
- For Statistics it is confirmed that the introduction of the selective test is associated with an improvement in the score.

Exploratory data analysis

Figure: Boxplot of ECTS for Courses of study, AER(OFA), and Gender



- Females have higher scores in both groups.
- In group "Additional Educational Requirements", the number of students is small for all courses of study.

Exploratory data analysis

- The distribution of the binary response variable Y:

Y	Category	%	F	M	Business Admin	Economics	Stats BD
1	≥ 40 ECTS	54.09	54.19	41.43	44.31	48.31	44.21
0	< 40 ECTS	45.91	45.81	58.57	55.69	51.69	55.79
			35.11	64.89	44.06	40.40	15.54

- The distribution of the ordinal response variable Y:

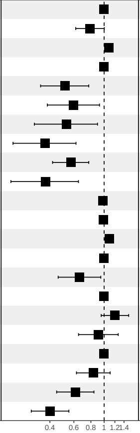
Y	Category	%	F	M	Business Admin	Economics	Stats BD
1	0 ECTS	17.23	10.87	20.67	18.56	15.65	17.54
2	from 1 to 39	36.86	34.94	37.90	37.13	36.03	38.25
3	from 40 to 59	26.94	29.04	25.80	22.15	31.44	28.77
4	60 ECTS	18.97	25.16	15.63	22.15	16.87	15.44
			35.11	64.89	44.06	40.40	15.54

Structure of the talk

Results of data analysis

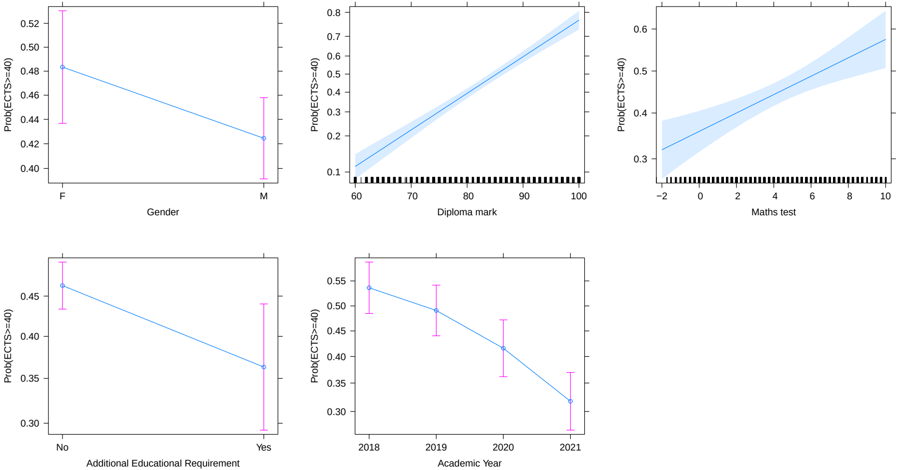
Figure: Forest plot for logistic regression

Variable		N	Odds ratio		p
Gender	F	644		Reference	
	M	1186		0.79 (0.62, 1.00)	0.053
DiplomaMark		1830		1.09 (1.07, 1.10)	<0.001
HighSchool	Scientific	913		Reference	
	Other	158		0.52 (0.34, 0.77)	0.001
	Classic	122		0.60 (0.38, 0.93)	0.021
	Linguistic	85		0.53 (0.31, 0.89)	0.019
	Professional	92		0.37 (0.21, 0.62)	<0.001
	TechnicalComm	374		0.57 (0.42, 0.77)	<0.001
	TechnicalInd	86		0.37 (0.21, 0.65)	<0.001
ScoreReading		1830		0.98 (0.93, 1.04)	0.538
ScoreLogic		1830		0.99 (0.94, 1.04)	0.668
ScoreMaths		1830		1.09 (1.05, 1.14)	<0.001
AER_OFA	No	1506		Reference	
	Yes	324		0.66 (0.46, 0.95)	0.024
CourseStudy	Business Admnistration	806		Reference	
	Economics	740		1.20 (0.95, 1.52)	0.122
	Stats for big data	284		0.91 (0.65, 1.27)	0.573
Year	2018	485		Reference	
	2019	519		0.83 (0.63, 1.11)	0.210
	2020	415		0.62 (0.45, 0.84)	0.003
	2021	411		0.40 (0.29, 0.55)	<0.001



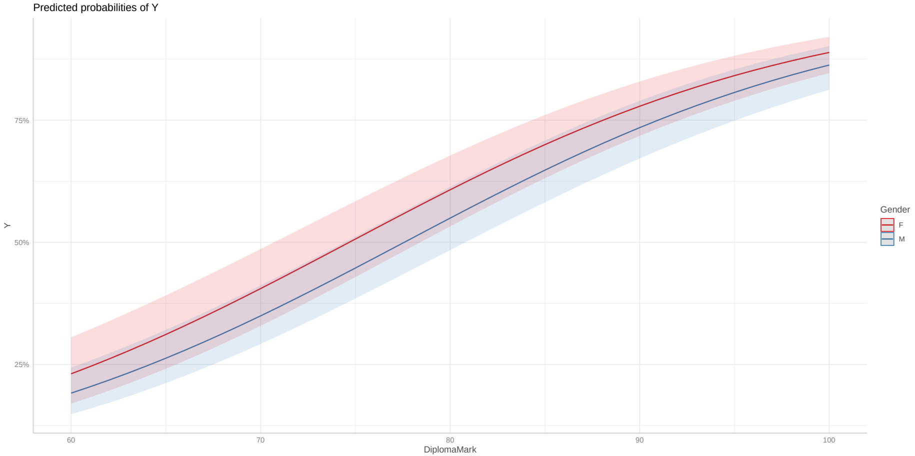
Results of data analysis

Figure: Effect plots for logistic regression - significant variables



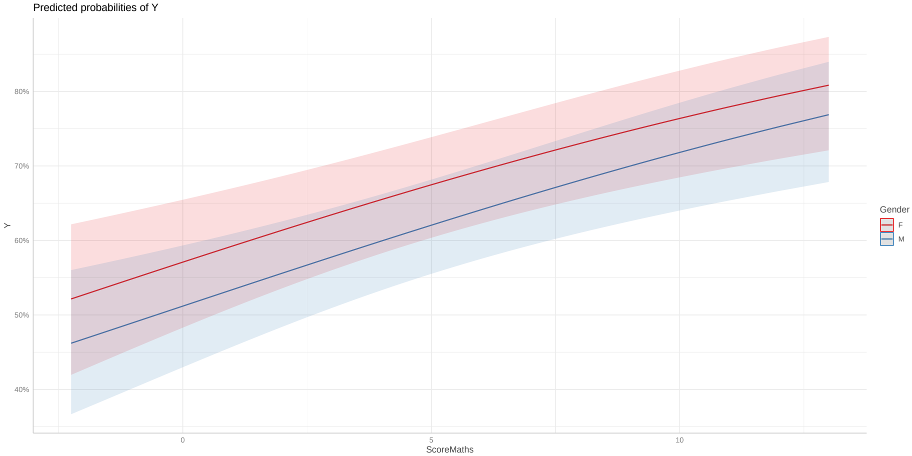
Results of data analysis

Figure: Effect plots for logistic regression - Diploma mark vs Gender



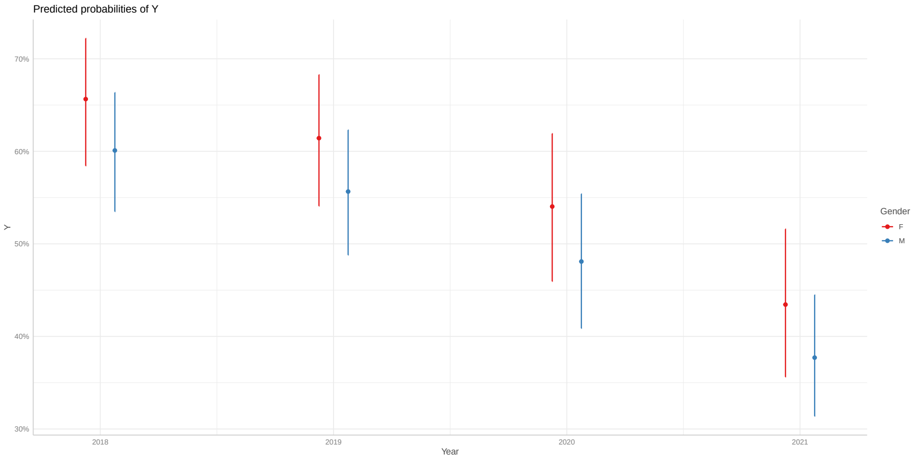
Results of data analysis

Figure: Effect plots for logistic regression - Maths Score vs. Gender



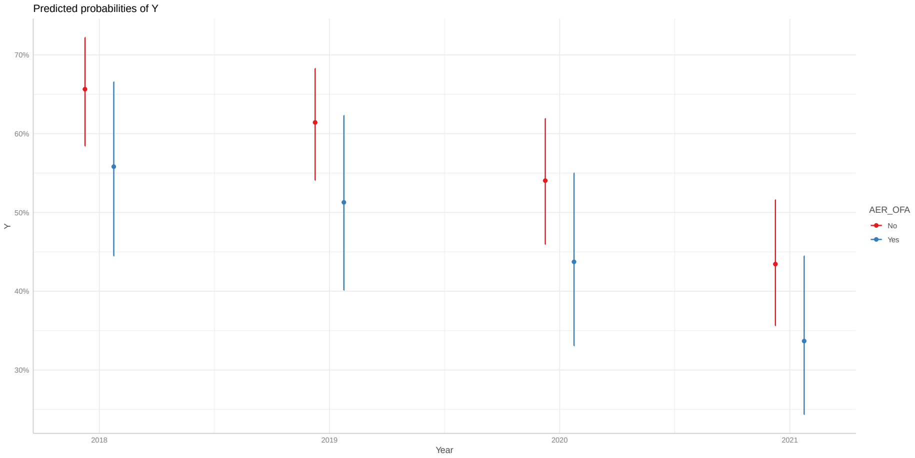
Results of data analysis

Figure: Effect plots for logistic regression - Academic year vs. Gender



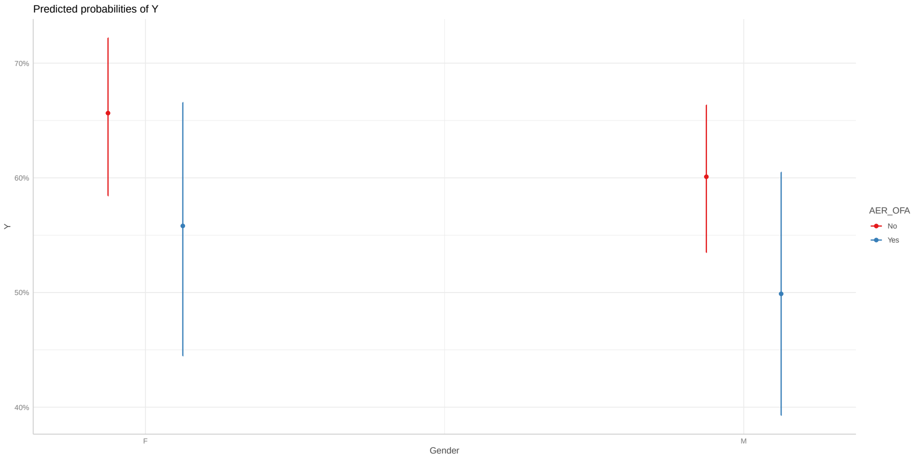
Results of data analysis

Figure: Effect plots for logistic regression - Academic year vs. AER (OFA)



Results of data analysis

Figure: Effect plots for logistic regression - Gender vs. AER (OFA)

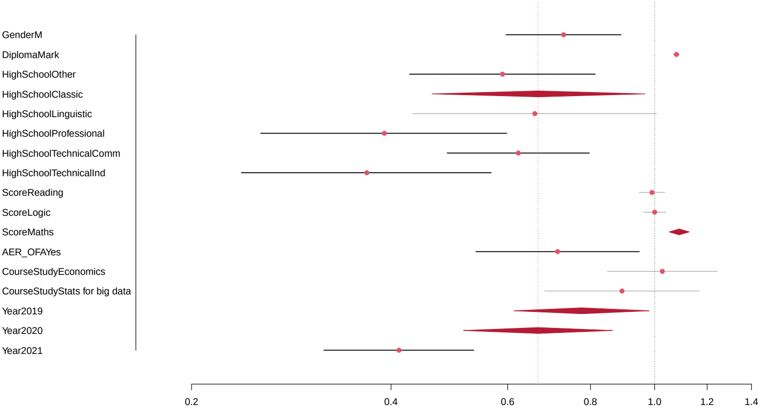


Go to Logistic regression plots.

Structure of the talk

Results of data analysis

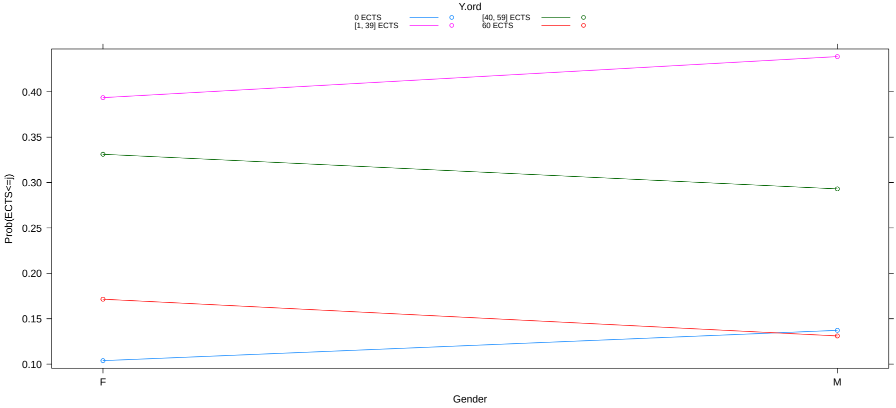
Figure: Forest plot for ordinal logistic regression



Variable	OR	95% CI	p-value
GenderM	0.73	[0.60–0.89]	0.0019
DiplomaMark	1.08	[1.07–1.09]	0
HighSchoolOther	0.59	[0.43–0.81]	0.0013
HighSchoolClassic	0.67	[0.46–0.97]	0.0325
HighSchoolLinguistic	0.66	[0.43–1.01]	0.0536
HighSchoolProfessional	0.39	[0.25–0.60]	0
HighSchoolTechnicalComm	0.62	[0.49–0.80]	2e−04
HighSchoolTechnicalInd	0.37	[0.24–0.57]	0
ScoreReading	0.99	[0.95–1.04]	0.6749
ScoreLogic	1.00	[0.96–1.04]	0.9998
ScoreMaths	1.09	[1.05–1.13]	0
AER_OFAYes	0.71	[0.54–0.95]	0.0199
CourseStudyEconomics	1.03	[0.85–1.24]	0.7842
CourseStudyStats for big data	0.89	[0.68–1.17]	0.4063
Year2019	0.78	[0.61–0.98]	0.0338
Year2020	0.67	[0.51–0.86]	0.0022
Year2021	0.41	[0.32–0.53]	0

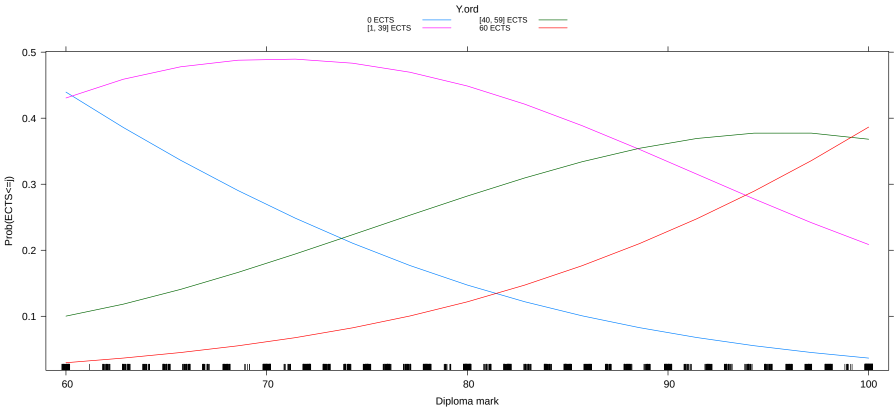
Results of data analysis

Figure: Effect for ordinal logistic regression - Gender



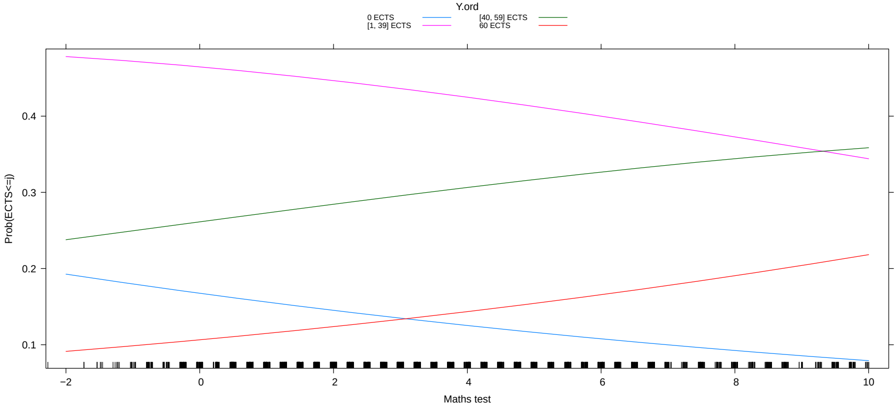
Results of data analysis

Figure: Effect for ordinal logistic regression - Diploma Mark



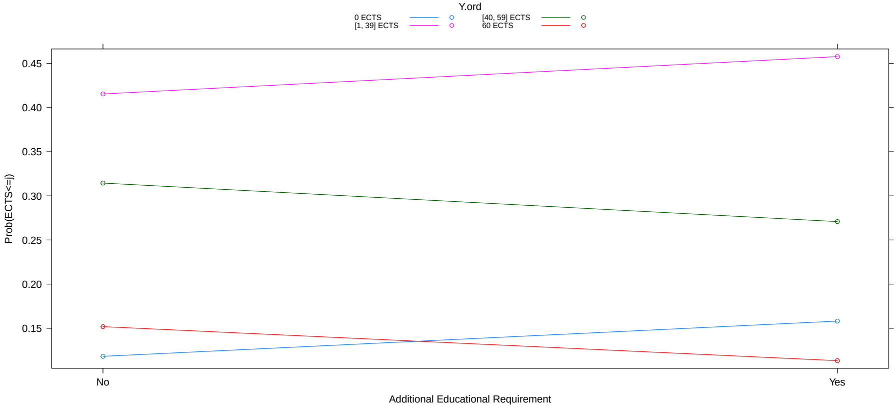
Results of data analysis

Figure: Effect for ordinal logistic regression - Maths test score



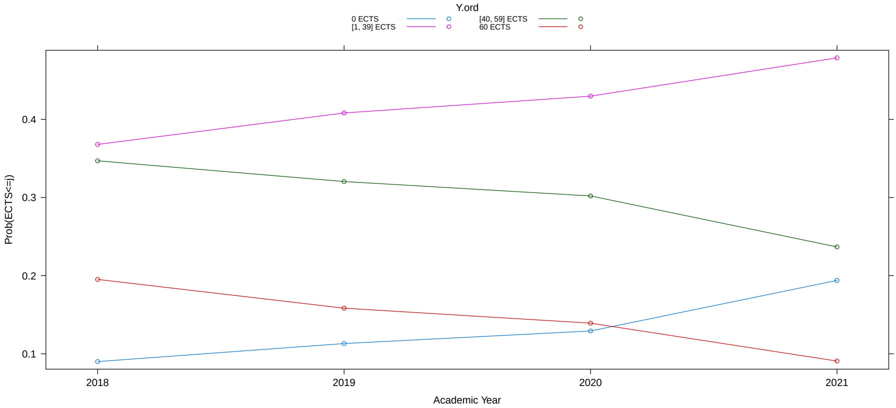
Results of data analysis

Figure: Effect for ordinal logistic regression - Additional Educational Requirement



Results of data analysis

Figure: Effect for ordinal logistic regression - Academic Year



Go to Ordinal logistic regression plots.

Structure of the talk

Conclusions

- We analyzed the performance of students in terms of ECTS earned during the first year.
- We focused on reaching the threshold of 40 ECTS.
- Two models have been considered.
    1. Logistic regression;
    2. Ordinal Logistic regression.

Conclusions

- The main research questions were:
  a) Does the pre-enrollment test score affect the probability of reaching the 40 ECTS during the first academic year?

     $\Longrightarrow$ Yes, only the maths test. A higher mark in maths is associated with an increase in the probability of reaching 40 ECTS.

  b) Is a gender gap detectable?

     $\Longrightarrow$ Yes, even if it is not particularly evident. Men have a lower probability of having more than 40 ECTS.

  c) Are there any differences among the courses of study?

     $\Longrightarrow$ No, there are not any statistically significant differences.

Conclusions

- Other comments:
  a) Having additional education requirements is associated with a lower probability of reaching 40 ECTS.
  b) An academic year trend is observed. The probability of having more than 40 ECTS decreases during the observation period.

     $\Longrightarrow$ Covid Effect?

  c) The type of high school affects the possibility of having at least 40 ECTS. In particular, the scientific lyceum has a higher probability of 40 ECTS, with respect to all other types.

     $\Longrightarrow$ Is it true for all courses of study? and what about gender?


Merci beaucoup
Thank You
お疲れ様
Danke
Gracias
Grazie
谢谢你
Thanks
Dank u
Obrigado

Structure of the talk

References

G. Adelfio, G. Boscaino, and V. Capursi (2014)
A new indicator for higher education student performance
Higher Educ. 68(5) (2014), 653–668

M. Attanasio, G. Boscaino, V. Capursi, and A. Plaia (2013)
May the students' career performance helpful in predicting an increase in universities income?
In: Statistical Models for Data Analysis. Series in Studies in Classification, Data Analysis, and Knowledge Organization, P. Giudici, S. Ingrassia, and M. Vichi, eds., Springer International Publishing, Switzerland, 2013.

G. Boscaino, V. Capursi, and F. Giambona (2007)
The careers' performance of a University students' cohort.
DSSM Working paper, n. 2007.1

References

L. Grilli, C. Rampichini, and R. Varriale (2013)
Predicting students academic performance: A challenging issue in statistical modelling.
In: Cladag 2013 Book of abstracts, Monerva, Morlini and Palumbo, eds., CLEUP, Padova, 2013, pp. 249–254.

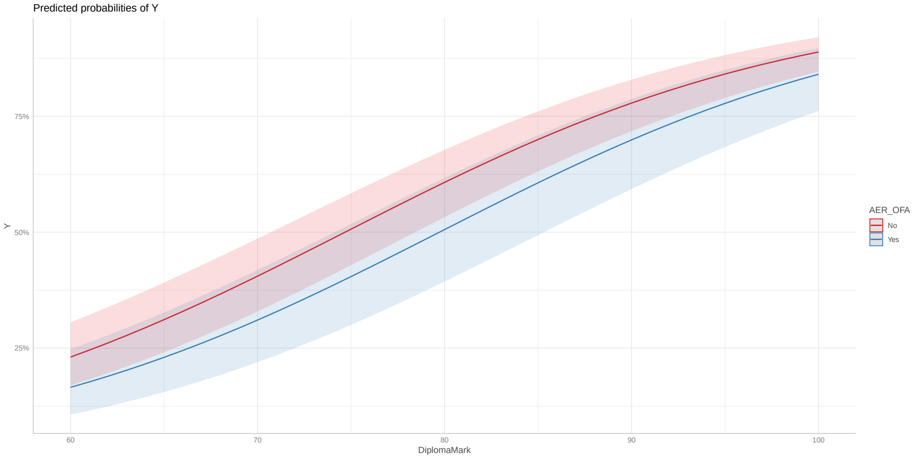
L., Grilli, C. Rampichini, R. Varriale (2015)
Statistical modelling of gained university credits to evaluate the role of pre-enrolment assessment tests: An approach based on quantile regression for counts.
Statistical Modelling. 16(1): 47–66.

Structure of the talk

Structure of the talk

Logistic regression plots

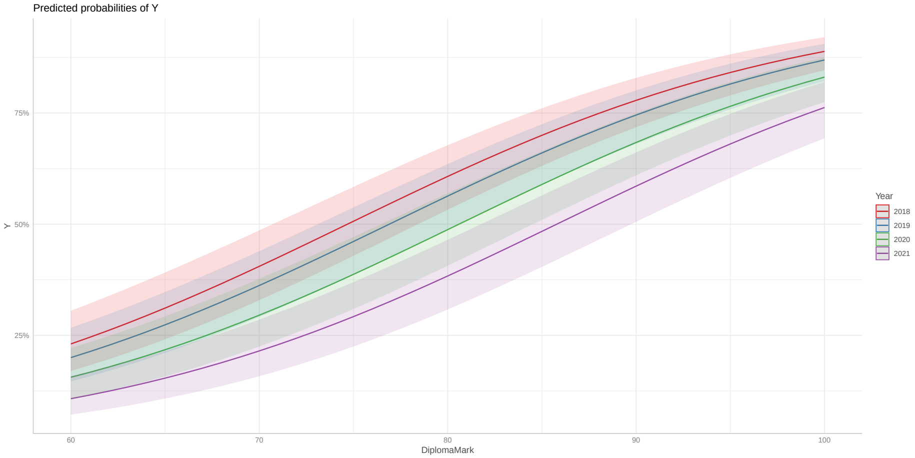
Figure: Effect plots for logistic regression - Diploma mark vs Additional Education Requirements



Go back to Results of logistic regression.

Logistic regression plots

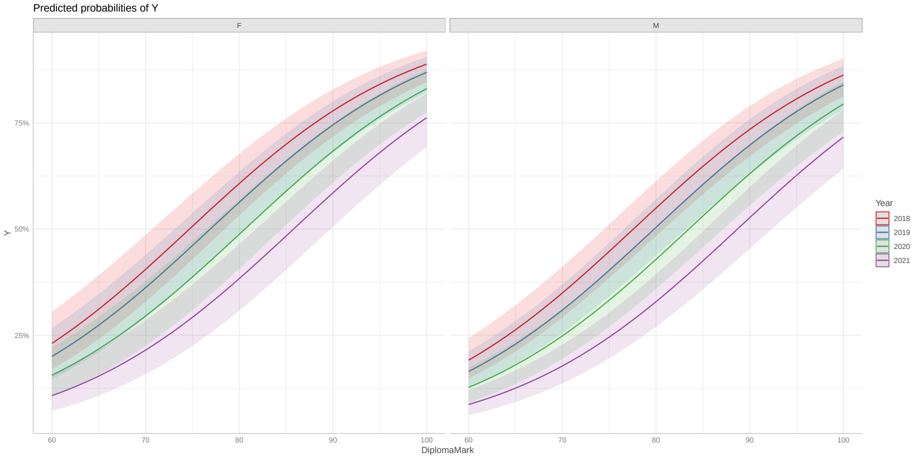
Figure: Effect plots for logistic regression - Diploma mark vs Academic Year



Go back to Results of logistic regression.

Logistic regression plots

Figure: Effect plots for logistic regression - Diploma mark vs Gender and Academic Year

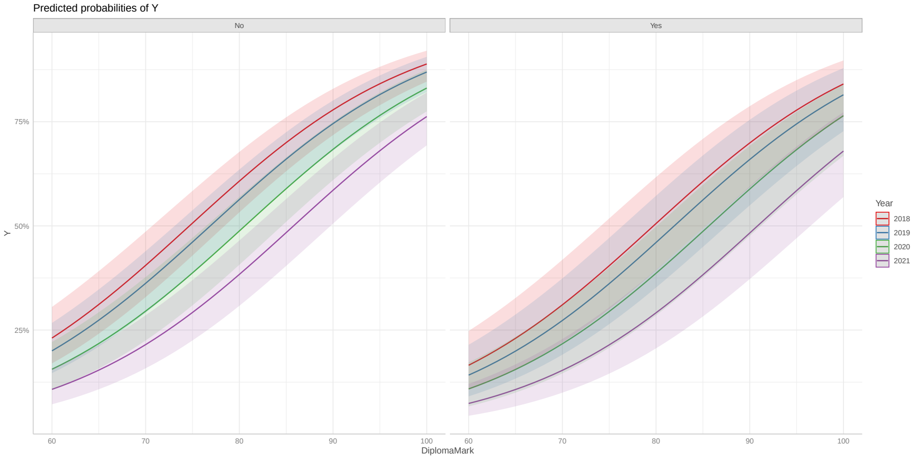


Go back to Results of logistic regression.

Go back to Results of logistic regression.

Logistic regression plots

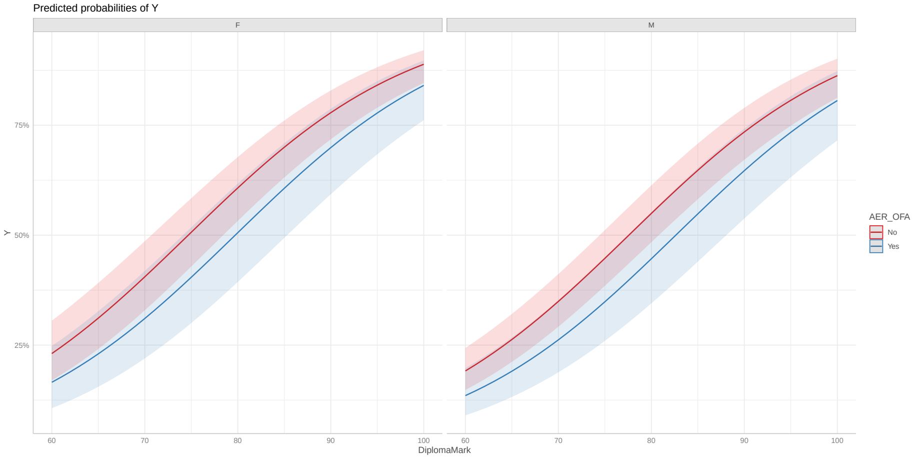
Figure: Effect plots for logistic regression - Diploma mark vs Additional Education Requirements and Academic Year



Go back to Results of logistic regression.

Logistic regression plots

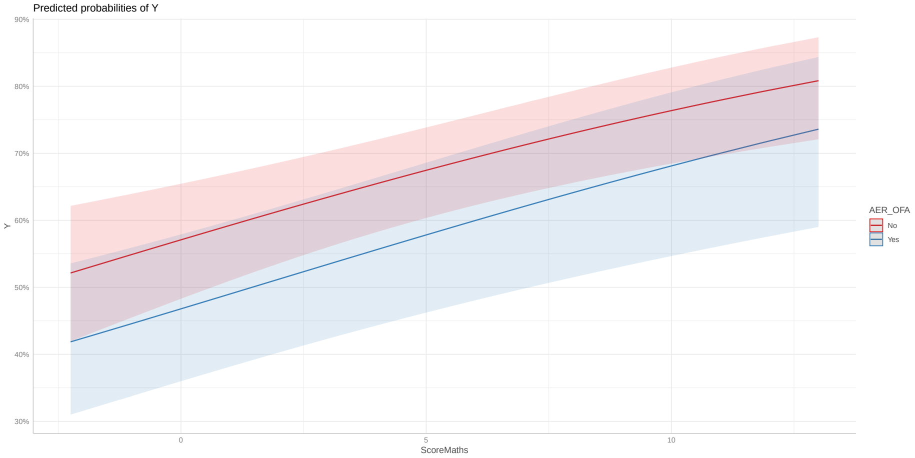
Figure: Effect plots for logistic regression - Diploma mark vs Gender and Additional Education Requirements



Go back to Results of logistic regression.

Logistic regression plots

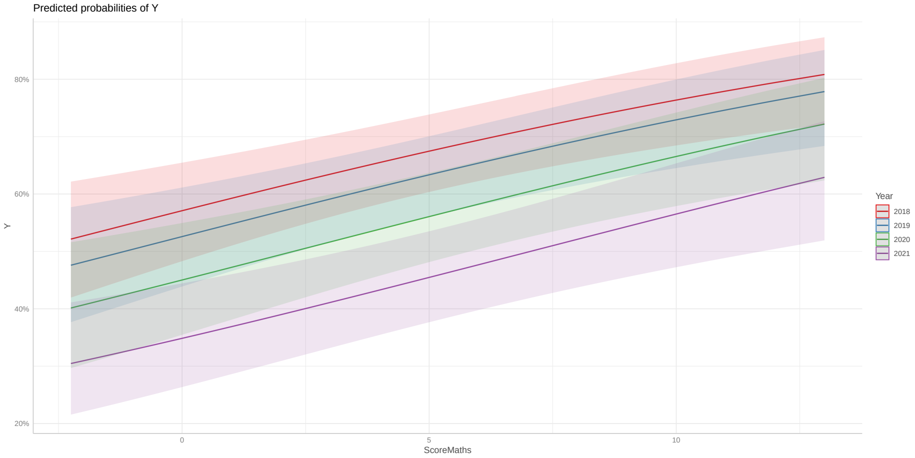
Figure: Effect plots for logistic regression - Maths Score vs Additional Education Requirements



Go back to Results of logistic regression.

Logistic regression plots

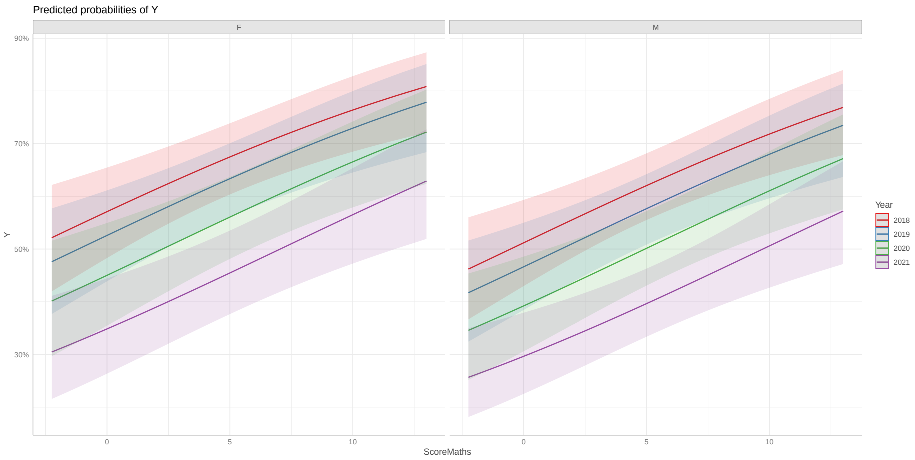
Figure: Effect plots for logistic regression - Maths Score vs Academic Year



Go back to Results of logistic regression.

Logistic regression plots

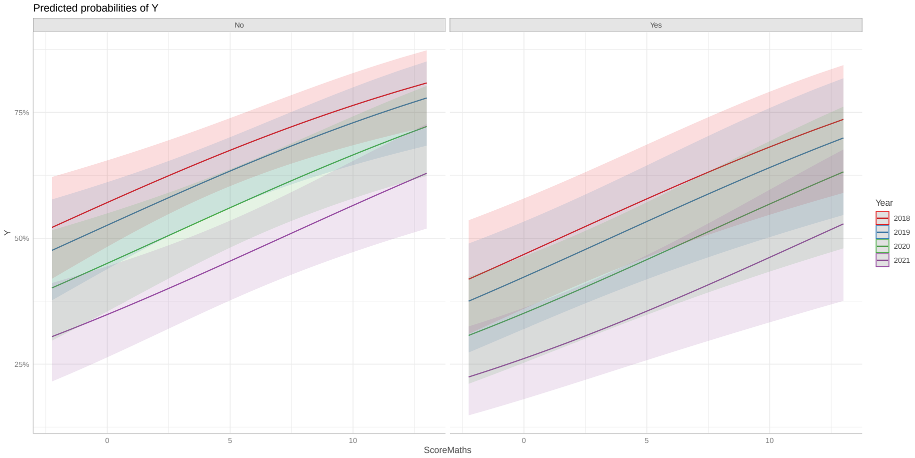
Figure: Effect plots for logistic regression - Maths Score vs Academic Year and Gender



Go back to Results of logistic regression.

Logistic regression plots

Figure: Effect plots for logistic regression - Diploma mark vs Academic Year and Additional Education Requirements



Go back to Results of logistic regression.

Logistic regression plots

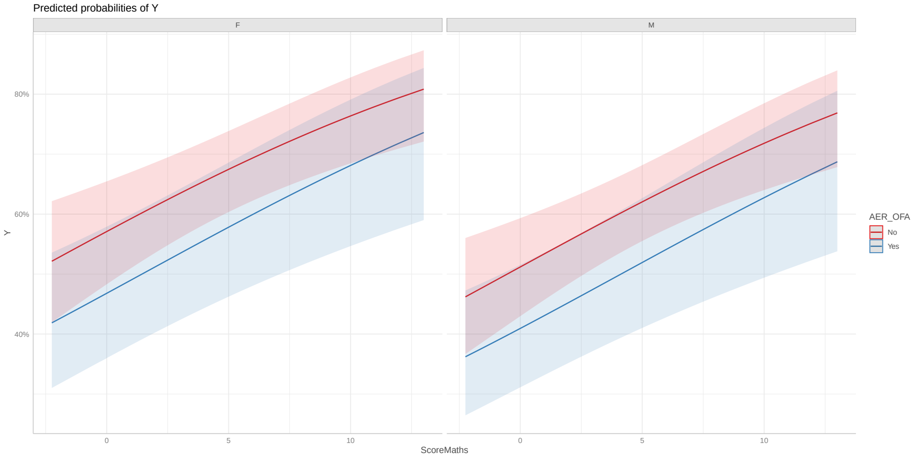
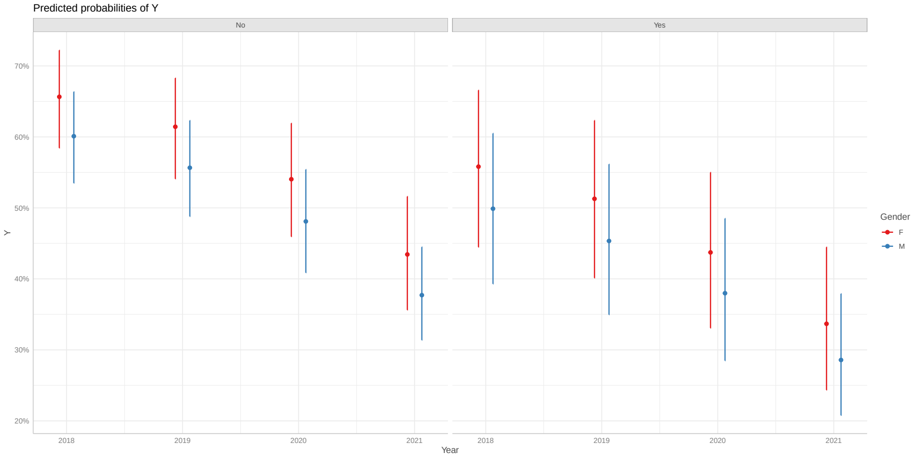


Figure: Effect plots for logistic regression - Maths Score vs Gender and Additional Education Requirements

Go back to Results of logistic regression.

Logistic regression plots

Figure: Effect plots for logistic regression - Academic Year vs Gender and Additional Education Requirements

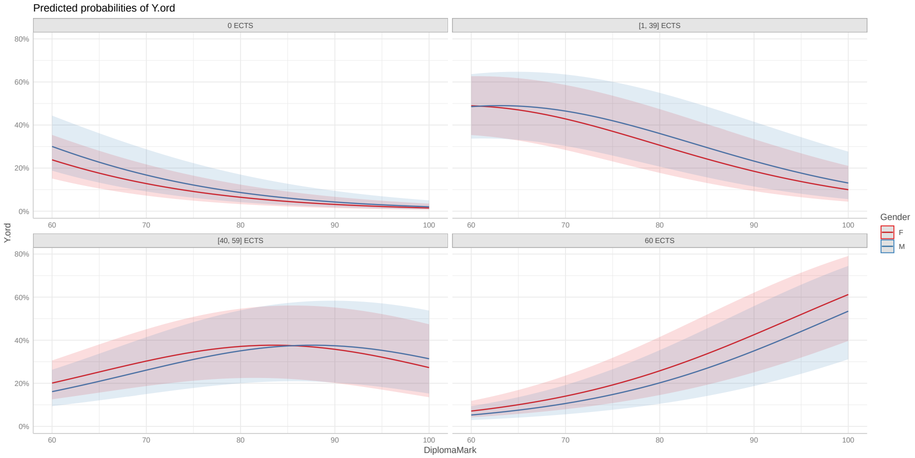


Go back to Results of logistic regression.

Structure of the talk

Ordinal logistic regression plots

Figure: Effect plots for ordinal logistic regression - Diploma mark vs Gender



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

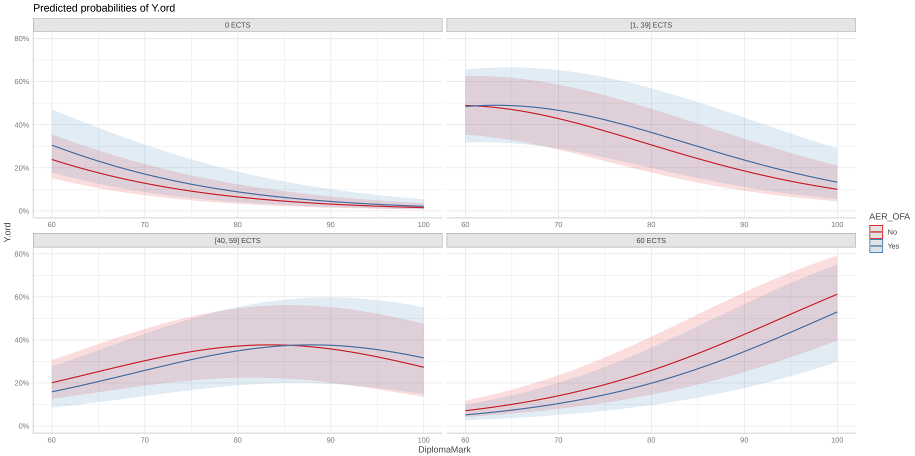
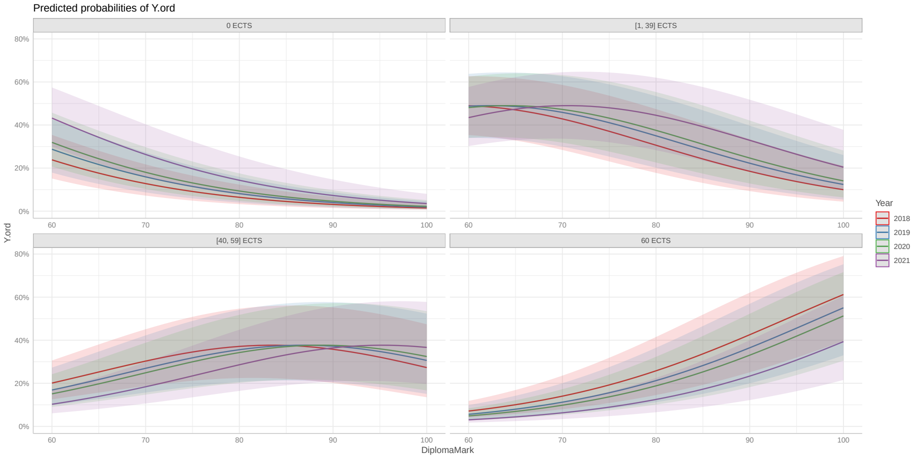


Figure: Effect plots for ordinal logistic regression - Diploma mark vs Additional Education Requirements

Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

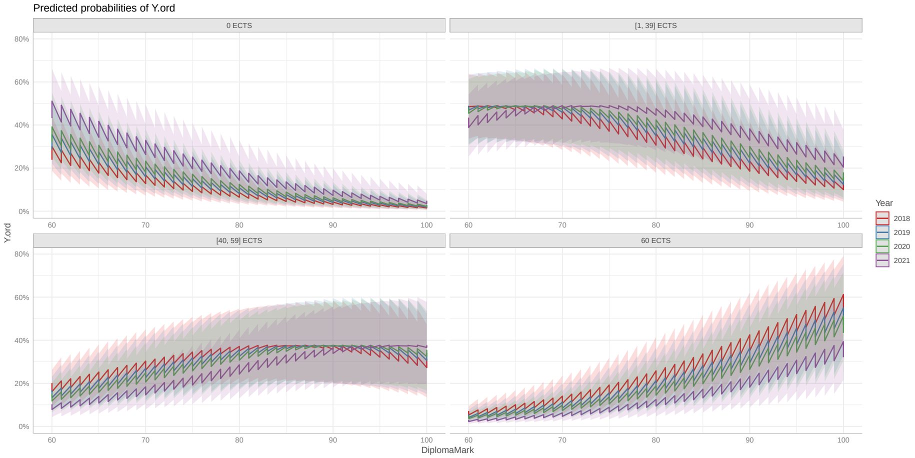
Figure: Effect plots for ordinal logistic regression - Diploma mark vs Academic Year



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

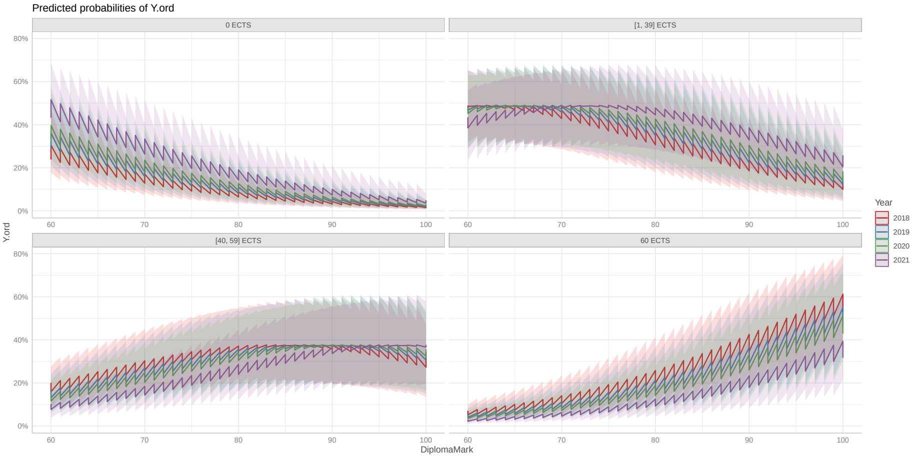
Figure: Effect plots for ordinal logistic regression - Diploma mark vs Academic Year and Gender



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

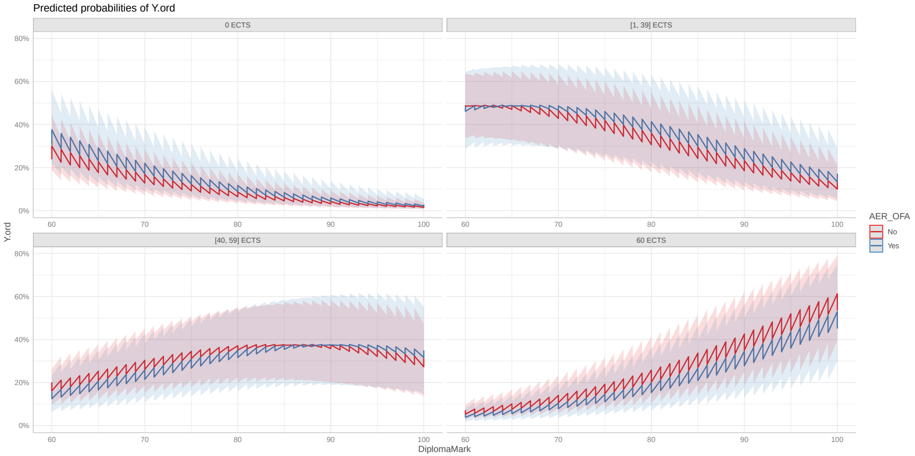
Figure: Effect plots for ordinal logistic regression - Diploma mark vs Academic Year and Additional Education Requirements



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

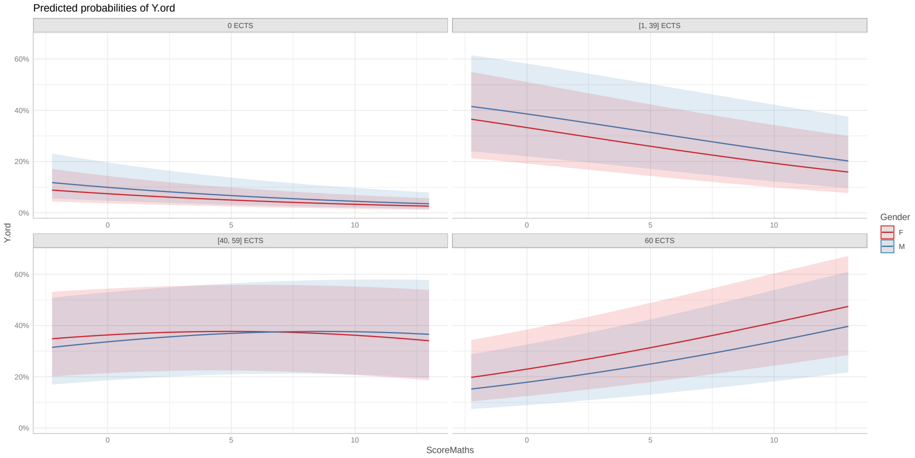
Figure: Effect plots for ordinal logistic regression - Diploma mark vs Additional education requirements and Gender



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

Figure: Effect plots for ordinal logistic regression - Maths Score vs Gender



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

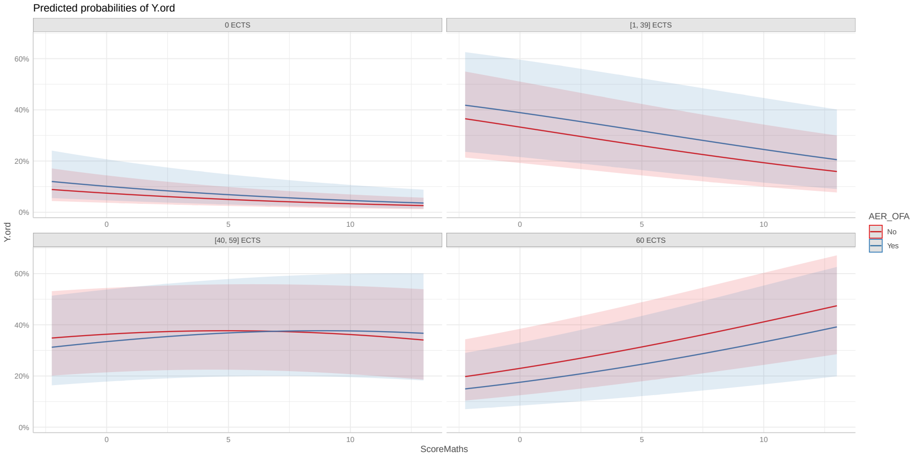
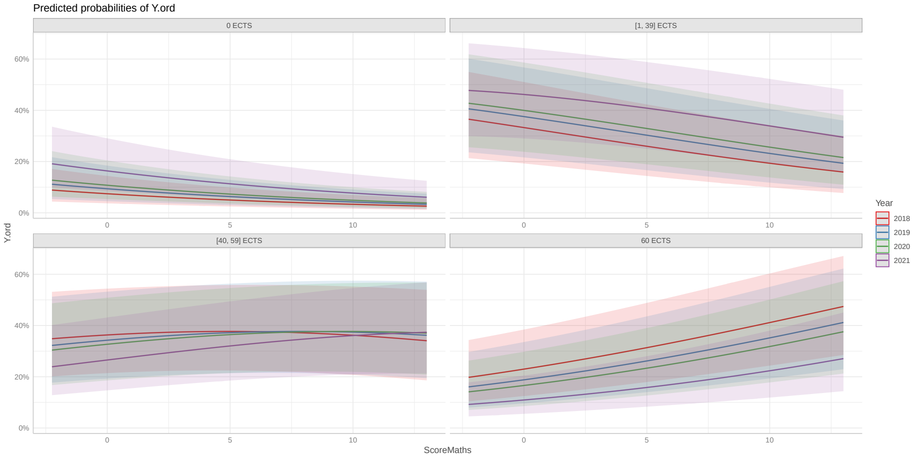


Figure: Effect plots for ordinal logistic regression - Maths Score vs Additional Education Requirements

Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

Figure: Effect plots for ordinal logistic regression - Maths Score vs Academic Year



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

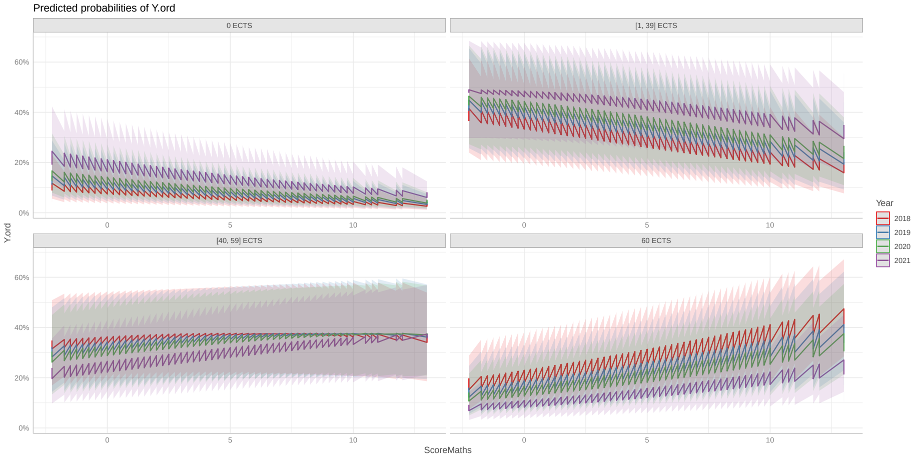
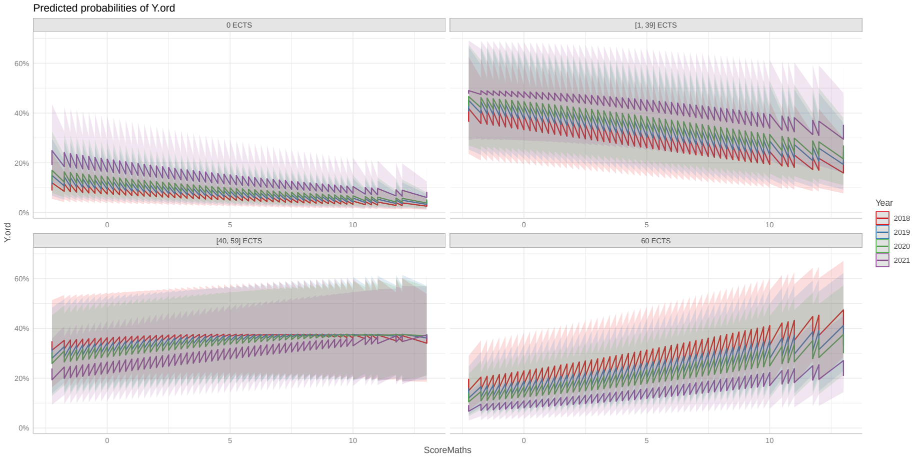


Figure: Effect plots for ordinal logistic regression - Maths Score vs Academic Year and Gender

Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

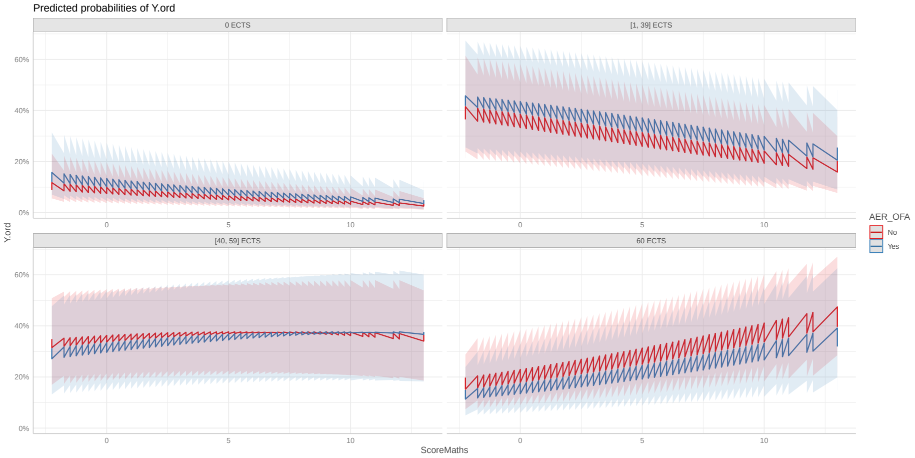
Figure: Effect plots for ordinal logistic regression - Maths Score vs Academic Year and Additional Education Requirements



⏩ Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

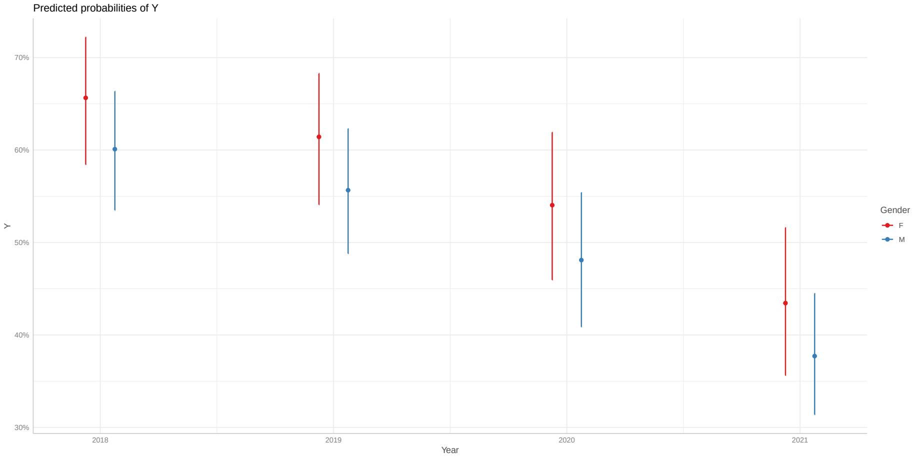
Figure: Effect plots for ordinal logistic regression - Maths Score vs Additional education requirements and Gender



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

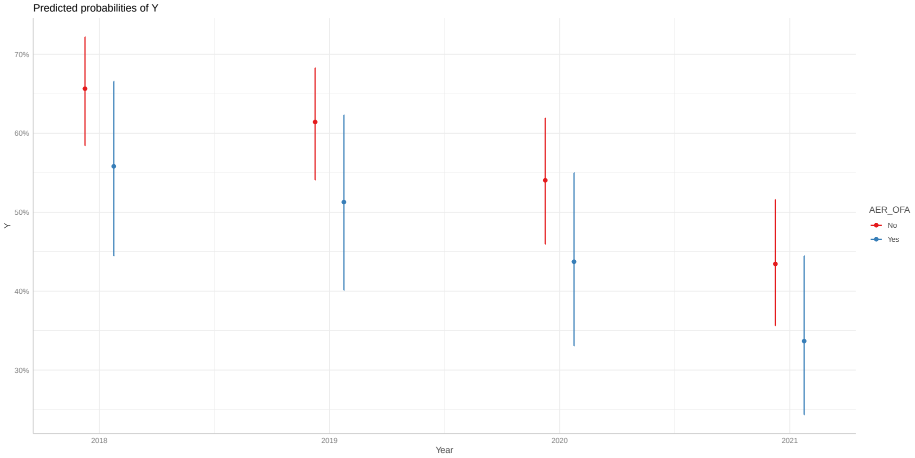
Figure: Effect plots for ordinal logistic regression - Academic Year vs Gender



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

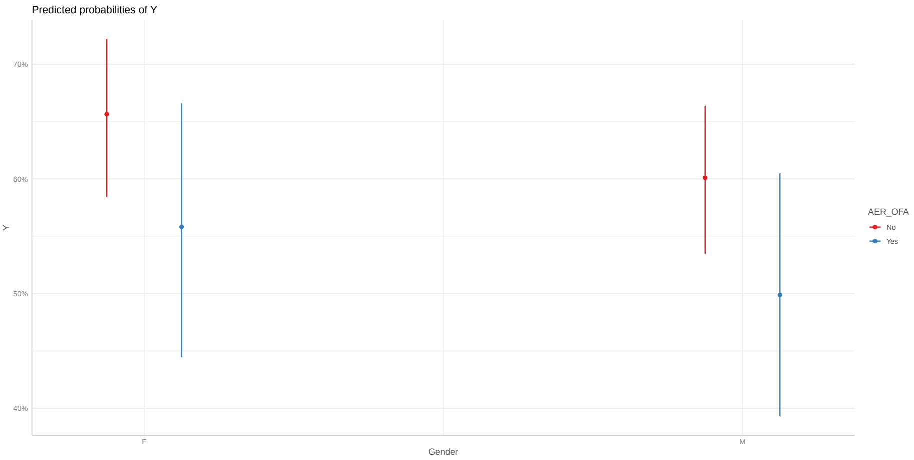
Figure: Effect plots for ordinal logistic regression - Academic Year vs Additional education requirements and Gender



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

Figure: Effect plots for ordinal logistic regression - Gender vs Additional education requirements



Go back to Results of ordinal logistic regression.

Ordinal logistic regression plots

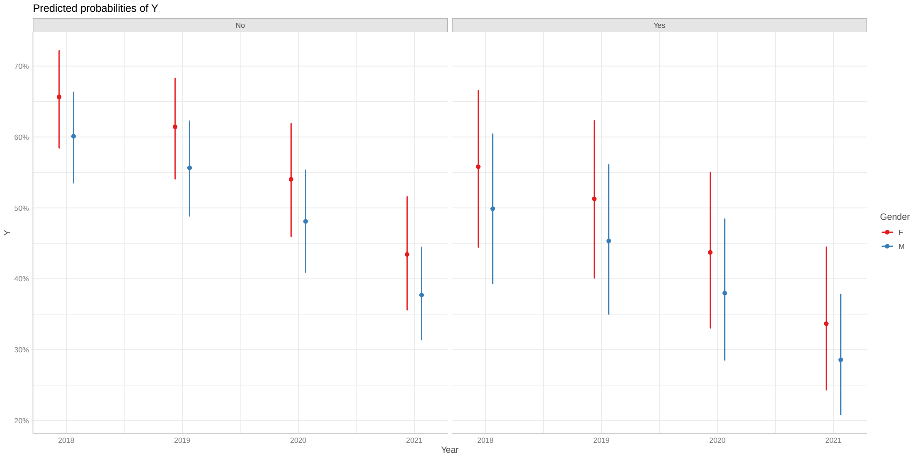


Figure: Effect plots for ordinal logistic regression - Academic Year vs Gender and Additional education requirements and Gender

Go back to Results of ordinal logistic regression.