

Undergraduate Statistics Admissions at the University of Warwick

RSS conference, Cardiff

Ella Kaye

supervised by Julia Brettschneider and Anastasia Papavasiliou

September 6th, 2018

The Admissions Process



Event tree for the admissions process.

Aims

- Model the university admissions process
- Build a decision tool for the University of Warwick Statistics department to use during their admissions process
 - 'best' students
 - right number of them

Aims

- Model the university admissions process
- Build a decision tool for the University of Warwick Statistics department to use during their admissions process
 - 'best' students
 - right number of them

Some sub-questions

- Who is likely to meet the offer (predicting A-level results)?
- Can we predict university performance from data available during the admissions period?

Data

5031 applicants for the 2011 - 2016 admissions cycles

5031 applicants for the 2011 - 2016 admissions cycles

1089 enrolled

5031 applicants for the 2011 - 2016 admissions cycles

1089 enrolled

1022 take first year exams

What predictors?

What predictors?

- **Academic**
 - Further Mathematics result (where available)
 - Best 'other' A-level result
 - Whether they have a Physics/Chemistry A-level
 - Number of A-level results
 - Number of A*/A grades at GCSE, for students with at least 8 GCSEs (imputed if missing)

What predictors?

- **Academic**

- Further Mathematics result (where available)
- Best 'other' A-level result
- Whether they have a Physics/Chemistry A-level
- Number of A-level results
- Number of A*/A grades at GCSE, for students with at least 8 GCSEs (imputed if missing)

- **Contextual**

- School Type (State/Grammar/Independent)
- POLAR3
- Index of Multiple Deprivation (IMD)

What predictors?

- **Academic**

- Further Mathematics result (where available)
- Best 'other' A-level result
- Whether they have a Physics/Chemistry A-level
- Number of A-level results
- Number of A*/A grades at GCSE, for students with at least 8 GCSEs (imputed if missing)

- **Contextual**

- School Type (State/Grammar/Independent)
- POLAR3
- Index of Multiple Deprivation (IMD)

- **Other**

- Gender
- Application period

Predicting A-levels

Making matches

- Data are really messy
 - Different ways of counting A-levels; hundreds of inconsistencies
 - **4469** results without predictions and **963** predictions without results
- Across all students and subjects, we can match **11530** predicted grades and results, for **3402** students.

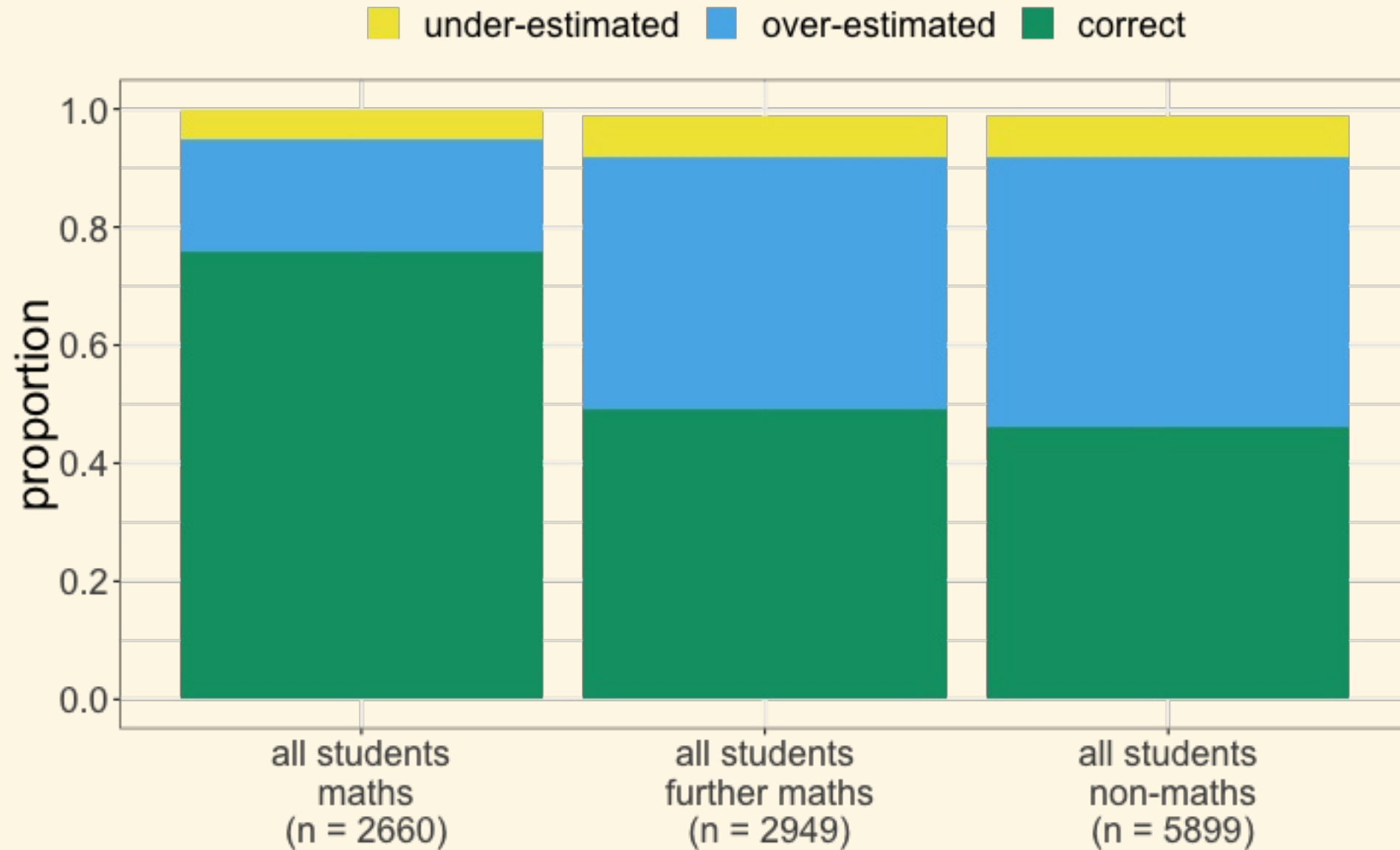
Predicted grades are

correct **54%** of the time

over-estimated **39%** of the time

under-estimated **7%** of the time

By subject



Can we do any better?

- Home students (those with home postcodes in England), who are taking at least 3 A-levels at a UK school (GCSE covariates imputed where necessary) **1320** of these.
- Train on 80% on data, test on 20%.
- Fit two logistic regression models for predicting A-level Mathematics result (A* or lower):
 - First with predicted grade as only covariate
 - Second with predicted grade plus school type, average GCSE score, number of A-levels, whether they're taking Physics/Chemistry A-level, Gender and Application period.

Can we do any better?

- Second model is a significantly better fit to the data on the training set.
- Teachers'/first model accuracy on test set is **70.1%**.
- Accuracy on test set for second model is **72.3%**.

Can we do any better?

- Second model is a significantly better fit to the data on the training set.
- Teachers'/first model accuracy on test set is **70.1%**.
- Accuracy on test set for second model is **72.3%**.
- Repeating for predicting A*/A at Further Maths, second model is a better fit on the training data, but not much in it on the test data (**76.2%** vs **77%**).

Predicting university outcomes

What outcomes?

- Drop-out / non-engagement (binary)
- Proceeding to the second year (binary)
- Resits required (binary)
- Good pass (binary)
- Classification (ordered categories)
- Year mark (percentage)

Which students (for now)?

- Home students (those with home postcodes in England), who are taking at least 3 A-levels at a UK school, who have achieved at least some marks towards at least some modules. There are **459** of these.
 - We do not know how well students who were not admitted to Warwick would have done had they had the opportunity.

First year degree classification

- Proportional odds logistic regression model:

$$\text{logit}[P(Y \leq j)] = \alpha_j - \beta x, \quad j = 1, \dots, J-1$$

- Covariates significant at 5%:

| Variable | OR | 2.5 % | 97.5 % |
|---|------|-------|--------|
| FM (A or lower) - relative to A* | 0.40 | 0.27 | 0.58 |
| FM (no result) - relative to A* | 0.32 | 0.10 | 0.98 |
| Best other A-level (A) - relative to A* | 0.41 | 0.26 | 0.62 |
| Best other A-level (lower) - relative to A* | 0.31 | 0.16 | 0.59 |
| Has Physics/Chemistry A-level | 2.05 | 1.36 | 3.10 |
| Num A*/A at GCSE | 1.19 | 1.08 | 1.31 |
| Male | 1.53 | 1.01 | 2.34 |
| Late application - relative to early | 0.26 | 0.08 | 0.79 |

Considerations

- Why are contextual factors that have found to be significant in the literature not significant here?
- What about likely important factors that we do not have in the admissions data (e.g. personality traits, study habits, what happens once they are at university)?

Thank you!

Any questions?

I'd love to hear from you!

E.Kaye.1@warwick.ac.uk

[@ellamkaye](https://twitter.com/ellamkaye)

ellakaye.rbind.io

github.com/EllaKaye