

# Data Mining Math Graduate School Admissions

Jianqiu Kong, Jiahao Zhu, A.J. Hildebrand (Advisor)

**I** University of Illinois at Urbana-Champaign  
 Joint Mathematics Meetings, San Diego, CA, January 10-13, 2018

## Motivating Questions

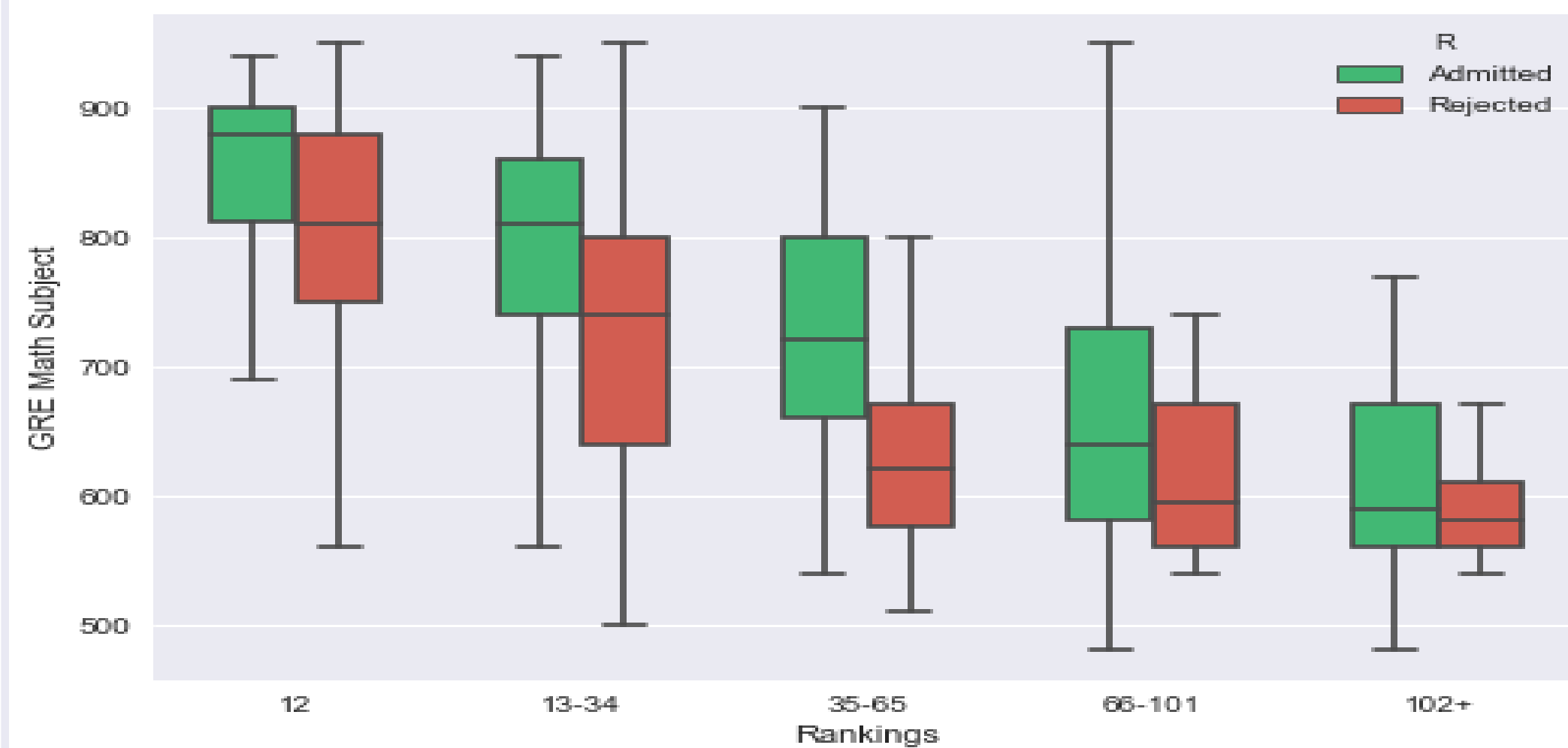
- How closely does the US News Ranking reflect the difficulty of getting accepted by a school? Is a higher ranking school always harder to get admitted to than a lower ranking school?
- How well can we predict admission outcomes based on the GRE and GPA Scores of an applicant, and the ranking of the school?
- Given a student's GRE and GPA Scores, can we predict the ranking of schools at which the student can expect to be admitted?

## Description of Data



- We scraped 358 applicant profiles posted to **mathematicsgre.com** from 2012 to 2017
- We removed applications to Masters Programs as well as PhD Programs in Statistics, leaving only applications to **Mathematics PhD Programs**.
- We also removed applicants with Masters Degrees and applications with incomplete GRE and GPA Scores. This leaves **228 applicant profiles**.
- Our analysis is based on the resulting **1331 applicant/school combinations**. Each of these combinations includes the following data:
  - GRE Quantitative, Verbal, Writing, Math Subject, and GPA
  - School applied to and US News Score of the school
  - Admission Result

## Sample Descriptive Statistics: GRE Math Subject Scores vs School Ranking



- GRE Math Subject Scores show a clear downward trend as the school ranking increases.
- GRE Verbal and Quantitative Scores and GPA show a similar downward trend. However, the GRE Writing Score shows a flat pattern.

## Ranking Schools Based on Admission Results

### Modeling Admission Results as Wins and Losses

- If an applicant is accepted by school  $i$  and rejected by school  $j$ , we consider this as a **game** in which **school  $i$  wins over school  $j$** .
- The number of wins of school  $i$  over school  $j$  is the number of applicants accepted by school  $j$  and rejected by school  $i$ .

### Ranking Methods

- **USNEWS:** US News & World Report Math Graduate Program Ranking (derived from a numerical score for each school).
- **WP (Winning Percentage)** (Barrow, 2013): Ranking based on the percentage of games won.
- **RPI (Rating Percentage Index)** (Barrow, 2013): A weighted version of the winning percentage that takes into account the winning percentage of the opponents.
- **K (Keener Method)** (Keener, 1993): Ranking based on an eigenvector of a matrix associated with the game results.
- **BT (Bradley-Terry Model)** (Bradley & Terry, 1952): A method that assigns a winning probability to each pair  $(i, j)$  of teams of the form  $p_i/(p_i + p_j)$ . The numbers  $p_i$  are optimized using the maximum likelihood method.

### Results

University	USNEWS	WP	RPI	K	BT
MIT	1	2	3	3	2
Princeton	1	3	2	2	3
UCBerkeley	3	11	9	7	8
Harvard	3	1	1	1	1
UChicago	5	6	6	6	6
Stanford	5	4	5	4	5
UCLA	7	10	11	9	9
Caltech	7	7	8	10	7
Michigan	9	20	12	13	13
Columbia	9	5	4	5	4
Yale	9	17	7	8	10
NYU	9	35	20	17	28
Cornell	13	21	15	12	14
Wisconsin	14	28	26	29	26
Brown	14	12	10	11	11
Texas	14	18	24	26	20
UIUC	17	19	25	16	18
Northwestern	17	24	14	15	16
MarylandCollegePark	17	40	32	35	34
Minnesota	17	41	29	18	27
Pennsylvania	17	45	13	32	24
Duke	17	15	28	19	21
Rutgers	23	32	17	14	23
UCSD	23	16	19	20	17

- **USNEWS vs other Rankings:** The WP, RPI, K, BT rankings are based on actual admission data, therefore are better indicators of the chances of acceptance by a particular school than the rankings by US News.

## Comparison of Ranking Methods

How close are the rankings generated by our methods (WP, RPI, K and BT) to the US News Ranking? We measure the distance to the US News Ranking using the following methods:

- **SpearmanWeighted** (Langville & Meyer, 2012): A weighted version of the Spearman Footrule (Spearman, 1906), which defines the distance between two permutations  $\pi_1$  and  $\pi_2$  by

$$\|\pi_1 - \pi_2\| = \sum_{i=1}^n \frac{|\pi_1(i) - \pi_2(i)|}{\min(\pi_1(i), \pi_2(i))}$$

- **Kendall's Tau** (Kendall, 1948): The proportion of pairs  $(i, j)$  that are ranked differently in  $\pi_1$  and  $\pi_2$ .

Method	WP	RPI	K	BT
SpearmanWeighted	58.62	36.94	36.53	37.66
Kendall's Tau	0.27	0.18	0.24	0.19

- RPI gives the best match according to Kendall's Tau. K gives the best match according to SpearmanWeighted.
- WP gives the worst match according to both methods.

## Discussion

- The rankings derived from user-posted admission results are mostly consistent with the US News Rankings.
- Some schools (e.g., NYU and Maryland) are easier to get admitted to than their US News rank would predict.
- Among the numerical scores in our data, the GRE Math Subject is the most significant factor in admission outcomes, and the GRE Writing Score is the least significant.
- Among the machine learning methods we used to predict admission outcomes, Logistic Regression and SVM with Linear Kernel are most accurate.

## References

- Barrow, D, et al. (2013). *Ranking rankings: an empirical comparison of the predictive power of sports ranking methods*. Journal of Quantitative Analysis in Sports 9(2): 187-202.
- US News & World Report, Best Graduate Mathematics Programs. [www.usnews.com/best-graduate-schools](http://www.usnews.com/best-graduate-schools). Accessed 9/23/2017.
- Bradley, R. A. and M. E. Terry. (1952). *Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons*. Biometrika 39:324-345.
- Keener, J. P. (1993). *The Perron-Frobenius Theorem and the Ranking of Football Teams*. SIAM Review 35:80-93.
- Kendall, M. (1948). *Rank Correlation Methods*. Charles Griffin & Company Limited.
- Langville, A. N. & C. D. Meyer (2012). *Who's #1? : The Science of Rating and Ranking*. Princeton University Press.
- MathematicsGRE.com. [www.mathematicsgre.com](http://www.mathematicsgre.com). Accessed 9/23/2017.
- Spearman, C. (1906). 'Footrule' for measuring correlation. British Journal of Psychology, 2, 89-108.

## Predicting Admission Outcomes

### Multiple Linear Regression

- **Goal:** Predict the highest US News Score of a school at which a student can expect to be admitted.
- **Predictors:** GRE Verbal ( $V$ ), GRE Quantitative ( $Q$ ), GRE Writing ( $W$ ), GRE Math Subject ( $S$ ), GPA ( $G$ ).
- **Result:**

Predictors	S	G	V	W	Q
P-Value	$10^{-8}$	0.0067	0.0044	0.3307	0.4523

- **R-squared:** 0.3598
- **Regression Equation:**  

$$\text{Score} = 0.025Q + 0.003S + 0.6938G$$

- **Remarks:** The most significant predictor is the GRE Math Subject Score. GRE Quantitative and Writing Scores are insignificant.

### Machine Learning Methods

- **Goal:** Predict whether the student will be accepted or rejected at a particular school.
- **Predictors:** GRE Verbal, GRE Quantitative, GRE Writing, GRE Math Subject, GPA, US News Score of the School.
- **Validation:** 5-fold Cross Validation.

Method	Accuracy
SVM, Linear Kernel	0.758
SVM, RBF Kernel	0.669
Decision Tree, Depth 3	0.707
Decision Tree, Depth 4	0.721
Decision Tree, Depth 5	0.732
Logistic Regression	0.745