Statistical Analysis of the International Mathematical Olympiad

ARTHUR BERG

Every year, participating countries send six of their best high-school math students to compete in the prestigious International Mathematical Olympiad (IMO). The selection process varies for each country, but students will typically participate in several rounds of competition with increasing difficulty, with many countries sending their highest-scoring students to intensive training camps. Presently, over 100 countries participate, and over the past sixty years, there have been over 18,000 contestants attending this prestigious competition. A massive number of scoring results are publicly available, and we analyze this data, elucidating a unique perspective on the world's toughest math competition. Some of the key analyses in this paper include:

- identifying some of the most difficult and least difficult problems on the IMO;
- identifying countries that have the best performance on the IMO and exploring their performances over time;
- showing how certain countries do much better or worse on certain problem types (algebra, combinatorics, number theory, and geometry);
- demonstrating the home-field advantage of hosting the IMO; and
- exploring gender differences in participation rates by country and performances by problem type.

The Data

The official IMO website[1]provides data on each contestant's gender, country, and score on each of the six contest problems. In addition, the name of the host country for each year as well as the country of origination of each contest problem was collected from published IMO shortlists and the IMO Compendium book.[2] Additionally, each contest problem has been classified into one of four groups: algebra, combinatorics, geometry, or number theory. There are many contest problems that span multiple classifications, and for those problems, the classification into a single category can be somewhat subjective. Nevertheless, many of the officially released shortlisted problems provide official classifications of the problems into these four categories. The data from some of the early years of the competition are rather incomplete, and those years were excluded from the analysis, leaving 45 years of nearly complete scoring data.

In Figure 1(A), we see that the number of IMO participants has dramatically increased over time. Consistent with the increase in number of participating countries is a decrease in the average scores, as shown in Figure 1(B). Additionally, there is substantial variation in the scores each year, as depicted in Figure 1(C). In order to remove the variability in the year-by-year scoring, the scores are

---

[1] https://www.imo-official.org.

[2] Dušan Djukić, Vladimir Janković, Ivan Matić, and Nikola Petrović. *The IMO Compendium: A Collection of Problems Suggested for the International Mathematical Olympiads: 1959–2009*, 2nd ed., Springer, 2011.

https://doi.org/10.1007/s00283-020-10015-z

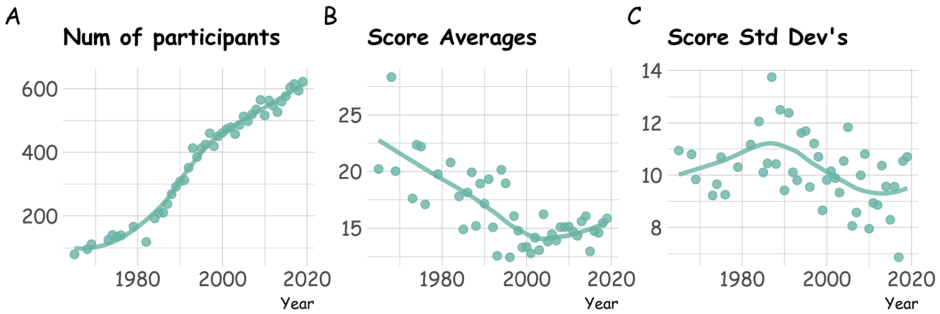

Figure 1. (A) Number of IMO participants by year. (B) Mean of the total scores (out of 42 points) by year. (C) Standard deviation of the total scores by year.

standardized in their respective year by subtracting that year's average and dividing by that year's standard deviation.

All of the raw data and R software code (tested on R version 3.6.3) used to generate the graphics, tables, and analyses included in this paper are available on GitHub.[3] The interested reader can easily reproduce and modify the analyses using the freely available and open-source software program R with various supporting packages.

The "Hardest" and "Easiest" IMO Problems

The IMO is a two-day contest in which students have 4.5 hours to solve three problems on each of the two days. By design, the first problem for each day (problems 1 and 4) are meant to be the easiest, the second problems (problems 2 and 5) are somewhat harder, and the last problems (problems 3 and 6) are intended to be the hardest.

There are a number of ways one might rank the many IMO problems over the years by difficulty. This analysis proceeds by ranking the problems based on standardized scores. After standardization, participant total scores in a given year have a mean of zero and a standard deviation of one. The top 10 easiest and hardest IMO problems, sorted by standardized average scores, are listed in Table 1. Based on this metric, the all-time hardest problem and easiest problem occurred in the same year and even on the same day: day 1 of the 2017 competition. Furthermore, day 2 of that 2017 competition was just as wild, with another extremely easy problem (relatively speaking) set alongside another extremely hard problem (all-time second-hardest problem).

Problems 1 and 3 of the 2017 contest are presented below to give the reader a sense of the difficulty of the all-time easiest and hardest problems. All previous contest problems are provided on the official IMO webpage,[4] which includes the option to download the problems in nearly sixty different languages. An expansive set of short-listed problems and solutions for the past several years is also available. Solutions to older contest problems can be found in the above-mentioned IMO Compendium book.

Table 1. Ten easiest and hardest IMO problems based on standardized average scores. The table includes year, problem number, problem category (algebra, combinatorics, geometry, or number theory), percentage of students who correctly solved the problem (scored 7 points for that problem), and standardized average score.

Easiest IMO Problems					Hardest IMO Problems				
Year	**#**	**Cat**	**%**	**Std Avg**	**Year**	**#**	**Cat**	**%**	**Std Avg**
2017	P1	N	72.6	0.51	2017	P3	C	0.3	−0.35
2006	P1	G	71.9	0.40	2017	P6	N	2.3	−0.31
2007	P4	G	69.8	0.39	2006	P6	G	1.6	−0.28
2017	P4	G	64.2	0.38	2010	P6	A	2.9	−0.27
2010	P1	A	57.0	0.37	2010	P3	A	3.1	−0.26
2012	P1	G	73.3	0.37	2007	P6	A	1.0	−0.25
2010	P4	G	70.7	0.35	2014	P6	C	2.7	−0.24
2011	P1	A	62.7	0.32	2011	P6	G	1.1	−0.24
2006	P4	N	49.8	0.32	2007	P3	C	0.4	−0.24
2015	P4	G	60.8	0.32	2012	P6	N	1.8	−0.23

IMO 2017, Problem 1 (all-time easiest) For each integer $a_0 > 1$, define the sequence $a_0, a_1, a_2, \ldots$ by:

$$a_{n+1} = \begin{cases} \sqrt{a_n} & \text{if } \sqrt{a_n} \text{ is an integer,} \\ a_n + 3 & \text{otherwise,} \end{cases} \qquad \text{for each } n \geq 0.$$

Determine all values of a_0 for which there is a number A such that $a_n = A$ for infinitely many values of n.

IMO 2017, Problem 3 (all-time hardest) A hunter and an invisible rabbit play a game in the Euclidean plane. The rabbit's starting point, A_0, and the hunter's starting point, B_0, are the same. After $n - 1$ rounds of the game, the rabbit is at point A_{n-1} and the hunter is at point B_{n-1}. In the nth round of the game, three things occur in order.

(i) The rabbit moves invisibly to a point A_n such that the distance between A_{n-1} and A_n is exactly 1.
(ii) A tracking device reports a point P_n to the hunter. The only guarantee provided by the tracking device is that the distance between P_n and A_n is at most 1.
(iii) The hunter moves visibly to a point B_n such that the distance between B_{n-1} and B_n is exactly 1.

---

[3] At https://github.com/arthurberg/IMO.

[4] https://www.imo-official.org/problems.aspx.

Is it always possible, no matter how the rabbit moves, and no matter what points are reported by the tracking device, for the hunter to choose her moves so that after 10^9 rounds she can ensure that the distance between her and the rabbit is at most 100?

Analysis by Country

Now we seek to determine the countries with the best overall scores. Recently, each country has been allowed to send up to six student participants to the IMO. In the early years of the IMO, teams consisted of eight participants from each country with a relatively small number of countries participating. This analysis considers only IMO results since 1984, when a six-member team became standard.

For each year, a country-specific standardized score is calculated by averaging the 36 standardized question scores for each country (six scores per participant, six participants per team). These country-, or team-based, scores are not calculated for any years in which a country may have sent fewer than six participants. These scores are then averaged over the years in which the respective teams participated in the IMO. The top 15 countries ranked by overall performance are listed in Table 2, which also includes the number of years that the country was considered in the analysis as well its current population. Data from the Soviet Union, which dissolved in 1991, has been excluded from this analysis. Similarly, data from East Germany, which reunited with Germany in 1990, is also excluded.

Table 2. The top 15 countries ranked by overall performance on the IMO.

	Std. Score	# of Years	Current Population (in millions)
China	11.32	33	1439
Russia	9.23	28	146
United States	9.18	36	331
South Korea	7.06	32	51
Romania	6.80	36	19
North Korea	6.37	12	26
Vietnam	6.33	36	97
Bulgaria	6.13	36	7
Taiwan	6.08	28	24
Hungary	5.70	36	10
Iran	5.58	33	84
Japan	5.53	30	126
Ukraine	4.97	27	44
Germany	4.93	36	84
United Kingdom	3.96	36	68

Table 2 shows a clear separation between the top three countries—China, Russia, and the United States—and the rest. It is also noted that Romania, Bulgaria, and Hungary all did quite well given that their populations are all under 20 million, which suggests a rich tradition of high-level mathematical problem-solving in those countries.

Figure 2 shows country-level performances by year for the top six countries. China has dominated the competition over the past 20 years, though in recent years, the USA has pulled ahead. South Korea has substantially improved its performance since 1990 and just barely missed first place in 2019, by only one point; the USA and China tied for first with 227 points each in this most recent contest. This record is particularly impressive when one considers that the populations of the USA and China are respectively more than 6 times and 26 times that of South Korea.

North Korea, ranked sixth with only 12 years of data, also demonstrated improved performance since 1990, but their record has been marred by various scandals. In particular, their scores were disqualified in 1991 and 2010 due to suspected cheating. More recently, in 2017 and 2018, North Korea did not participate in the IMO, which is presumably in response to the actions of one of their participants in 2016—Jong Yol-ri—who fled to the South Korean consulate while attending the IMO in Hong Kong.

The USA team's performance in 1994 is of particular interest, since all six members achieved a perfect score that year (hence yielding a team total of $6 \times 6 \times 7 = 252$ points). This is the only time all of a team's members achieved perfect scores, and it is unlikely to be repeated, though in 1987, Romania had a near-perfect team with a total score of 250. Even though the American team did so remarkably well in 1994, their performance does not stand out as spectacular in Figure 2, since the overall average that year was relatively high. In contrast, although the American team's score in 1996 was only 185 (they placed second that year, behind Romania), their relative performance that year was even higher than in 1994 due to the increased difficulty of the 1996 problems.

Analysis by Problem Type

Now we explore how countries performed relative to the four main problem types: algebra, combinatorics, geometry, and number theory. Figure 3(A) shows the distribution of standardized scores by problem type, which on the surface suggests that students tend to perform better on geometry problems; however, this is simply an artifact of Simpson's paradox. The mosaic plot in Figure 3 shows that in fact, geometry problems tend to occur more frequently among the easiest problems (problems 1 and 4) and less frequently among the hardest (problems 3 and 6). Similarly, though it is not quite as extreme, combinatorics problems tend to occur more frequently among the hardest problems and less frequently among the easiest problems.

In order to give each problem equal weighting, the scores have been standardized by year and problem number (as opposed to just year). This normalizes out the varying difficulty among the problems and allows us to explore whether certain countries do better or worse with certain problem types. As in the previous section, here we consider only IMO results since 1984, when a six-member team became standard.

We used analysis of variance to quantify the overall differences in performance of each country with respect to problem type. Based on this analysis, we report in Table 3 the 15 countries with the most widely differing scores

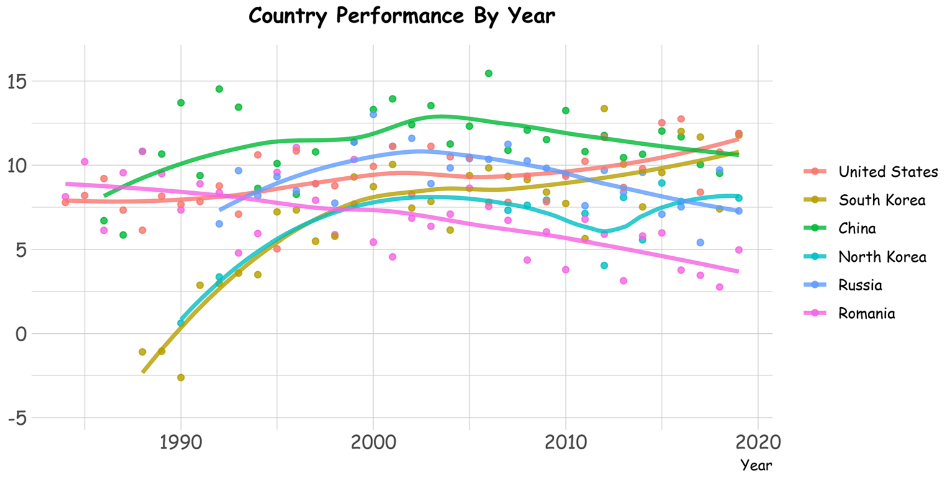

Figure 2. Country-level performances by year for the top six countries.

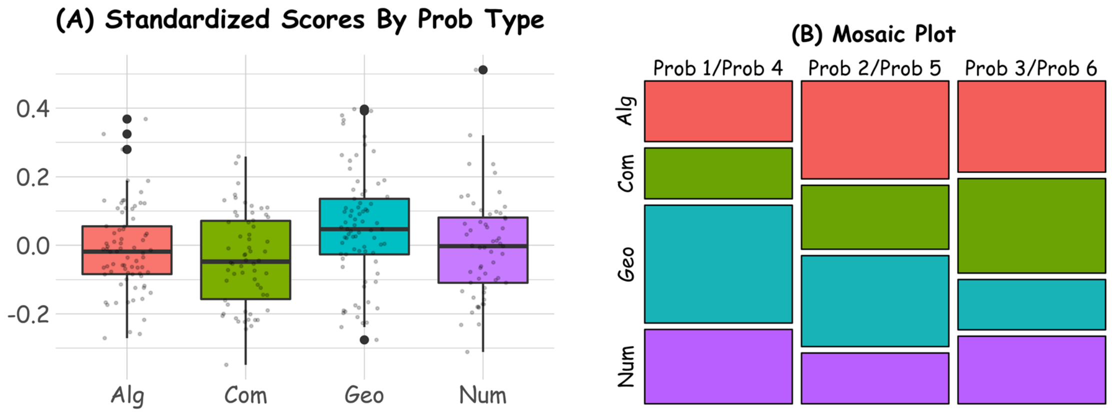

Figure 3. (A) Box plots and distributions of standardized scores by problem type showing slightly higher scores on geometry problems. (B) Mosaic plot showing that geometry problems occur more frequently as the first problem and less frequently as the last problem.

based on problem type. This table provides the average standardized score for each problem type and the number of problems included in calculating the average. Problem types with the highest and lowest averages are highlighted in italics and boldface, respectively, for each country. We notice the rather interesting phenomenon that countries tend to do either really well on geometry problems and poorly on combinatorics problems or the other way around.

To delve a little further, we highlight the profiles of three countries—Greece, Norway, and the United States of America—and visualize the strengths and weaknesses of these countries over time in Figure 4. Greece's performance on geometry problems is clearly much stronger than on the other problem types, which may be attributed to Greece's rich heritage in classical geometry having left a lasting influence on the country. Norwegian contestants, on the other hand, consistently perform the best on combinatorics problems. The USA contestants have historically performed rather poorly on geometry problems; in recent years, however, geometry has been their strength.

Home Advantage?

Each year, one of the participating countries serves as the host for the competition, and now we investigate whether there is a home advantage associated with being host. For this analysis, we keep the same standardization as in the previous section—scores are standardized within each year and problem number.

The data are aggregated into a data frame consisting of 2368 rows containing the year, country, standardized score, and a binary variable indicating whether the country in the given year was also the host country for that year. Here is the regression equation that was fit to the data with an adjusted R^2 of 0.85:

$$\text{Score} \sim \text{Host.Indicator} + \text{Country} * \text{Year} + \text{Country} * \text{Year}^2.$$

Based on this model, hosting the IMO was found to increase the host country's total score (out of 252 possible points), on average, by about 9.4 points, or about 9.8%; this effect is statistically significant, with a p-value of 0.0053.

Table 3. Top 15 countries with the most widely differing scores based on problem type. Average standardized scores for each problem type are followed by the number of problems included in calculating the average in parentheses. Problem types with the highest and lowest averages are highlighted in italics and boldface, respectively, for each country.

	Algebra	Combinatorics	Geometry	Number Theory	*p*-value
Greece	−0.39 (51)	−**0.56** (50)	*0.34* (60)	−0.40 (49)	4.7e−13
Norway	−0.66 (49)	*−0.18* (46)	−**0.92** (57)	−0.44 (46)	7.5e−11
Finland	−0.56 (53)	*−0.42* (52)	−**1.00** (62)	−0.55 (49)	2e−08
Azerbaijan	−0.49 (32)	−**1.00** (33)	*−0.18* (39)	−0.25 (28)	2.5e−08
Vietnam	1.20 (53)	**0.48** (52)	*1.30* (62)	1.00 (49)	4.1e−07
Tajikistan	−0.52 (16)	−**0.79** (18)	*0.32* (19)	−0.42 (13)	4.6e−07
Sweden	−0.240 (53)	*−0.049* (52)	−**0.650** (62)	−0.350 (49)	2.6e−06
Denmark	−0.51 (35)	*−0.18* (35)	−**0.86** (44)	−0.70 (30)	6.9e−06
Morocco	*−0.37* (45)	−**1.00** (44)	−0.62 (54)	−0.70 (43)	8.8e−06
Cuba	−**0.750** (21)	−0.710 (23)	*0.057* (23)	−0.540 (17)	1.5e−05
Uzbekistan	−0.35 (16)	−**0.99** (13)	*0.29* (19)	−0.38 (12)	2.7e−05
Iceland	−0.91 (33)	*−0.63* (36)	−**1.10** (43)	−0.96 (32)	6.1e−05
Peru	0.100 (28)	−**0.390** (33)	*0.490* (35)	0.039 (24)	6.9e−05
United Kingdom	0.75 (53)	*0.85* (52)	**0.30** (62)	0.77 (49)	0.00015
Romania	*1.30* (53)	**0.79** (52)	1.00 (62)	1.30 (49)	0.00025

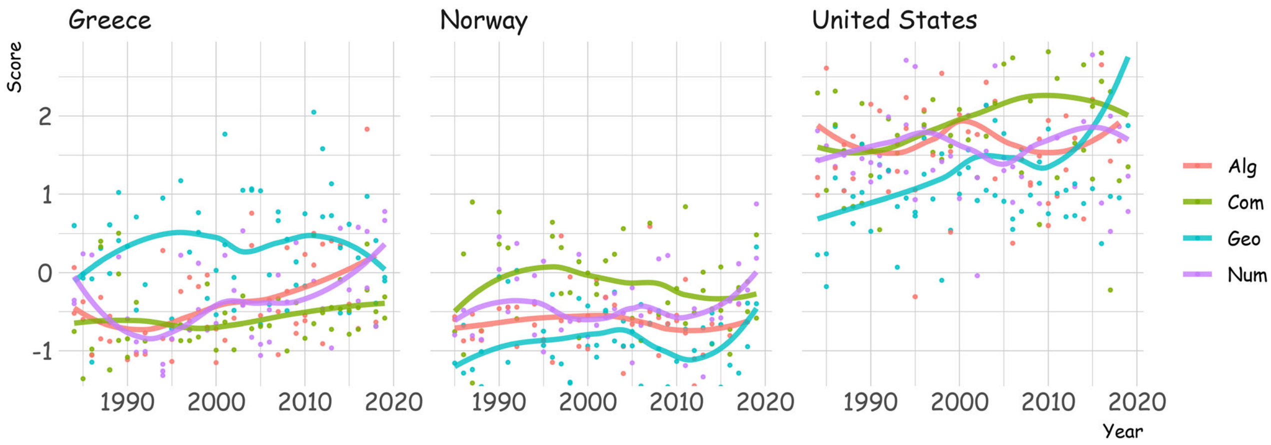

Figure 4. Performance over time by problem type for three selected countries.

Although statistically significant, there is still substantial variation, with the host country performing considerably better in some years and considerably worse in others. Figure 5 demonstrates this disparity with three selected countries—China, Russia, and the United States—with China having one of its best years as host, Russia having one if its worst years as host, and the United States having a better-than-average year as host.

We similarly explore the possibility of a "home advantage" with regard to the country of origination of the contest problems. That is, we consider whether a given country's team scores better on contest problems proposed by their country. The data have been aggregated into a data frame consisting of 14,208 rows containing the year, country, problem number, standardized score, and a binary variable indicating whether the problem was submitted by the country corresponding to that row. Here is the regression equation that was fit to the data with an adjusted R^2 of 0.57:

$$\begin{aligned}\text{Score} \sim \text{Submission.Indicator} + \text{Country} * \text{Year} \\ + \text{Country} * \text{Year}^2.\end{aligned} \tag{1}$$

Again, a statistically significant "home advantage" is found in that countries that provided the problem scored 16% better, on average, than what they would have scored otherwise; the *p*-value associated with this variable is 0.0004.

The data are too sparse to attempt to identify a significant home advantage for any particular host country; however, some countries have had several problems included on the IMO, which allows us to explore the effect of a home advantage for those individual countries. Only 14 countries (AUS, BGR, CZE, FRA, GBR, GER, IND, IRN, KOR, NET, POL, ROM, RUS, and USA) have had at least five problem submissions on the IMO, and we limit this analysis to just those countries. For each country, we assessed the problem contribution home advantage effect on each

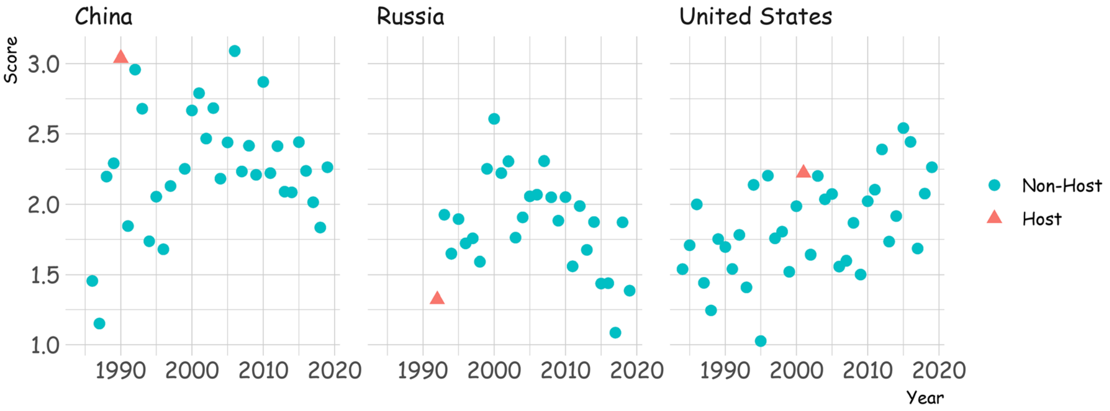

Figure 5. Being host is found, on average, to be associated with higher scores of the host country, but there is substantial variability within the home-turf advantage.

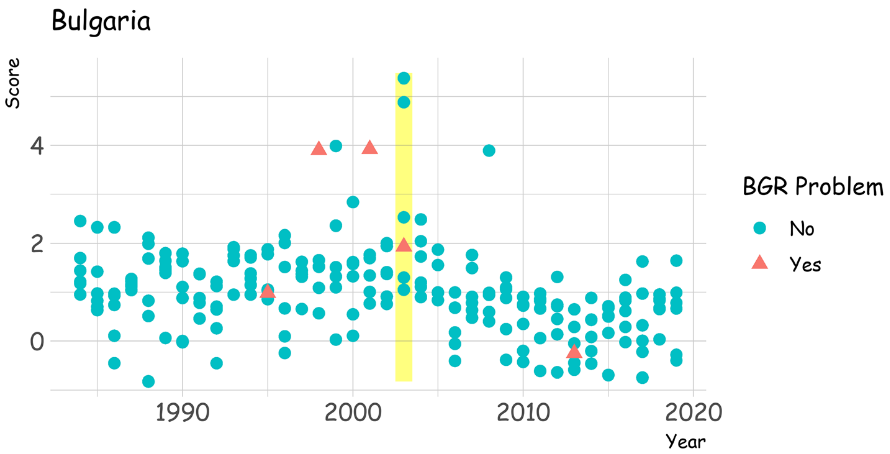

Figure 6. Bulgaria's performance on individual problems. Problems that Bulgaria contributed are differentiated. Bulgaria's first-place performance in 2003 is also highlighted.

country's score. Since we are conducting 14 tests, we impose a Bonferroni correction on the standard 5% type-1 error rate, which leads to $0.05/14 \approx 0.0036$ as the p-value threshold for statistical significance. Only one country—Bulgaria—was found to have a statistically significant home advantage, with a p-value of 0.0028; the second-smallest p-value is 0.093 (not even surpassing the unadjusted 0.05 level of significance). Figure 6 takes a closer look at Bulgaria's team scores. This figure demonstrates Bulgaria's exceptional performance on the two problems it contributed in 1998 and 2001 (the sixth problem for both years). Bulgaria also had a truly exceptional year in 2003 (highlighted in yellow), in which they placed first in the IMO. The second problem on that year's contest was contributed by Bulgaria.

Gender Differences

As a final analysis of the IMO data, gender patterns are explored. The proportion of female participants has been gradually increasing over the years, but unfortunately, current participation rates are only around 10%. If the mathematical sciences are to remain a vibrant choice for our students, the engagement of women at much higher rates is critical.

Figure 7 depicts the proportion of female participants by country. Poland (1.5%), Japan (1.7%), and the United States (2.0%) are among the countries with the lowest proportion of female participants. Countries with a long history of participation in the IMO and that have the highest proportions of female participants tend to be from southern and northern Europe, including North Macedonia (21%), Bosnia and Herzegovina (19.5%), Slovenia (17%), Latvia (15.2%), Croatia (14.2%), Ireland (13.5%), Estonia (13.5%), and Iceland (13.0%).

Figure 8 depicts standardized scores stratified by problem category and gender. These scores have been standardized for each problem so that individual problem difficulties have been normalized out. Male scores fall slightly above female scores for each category, but the difference is least pronounced for geometry.

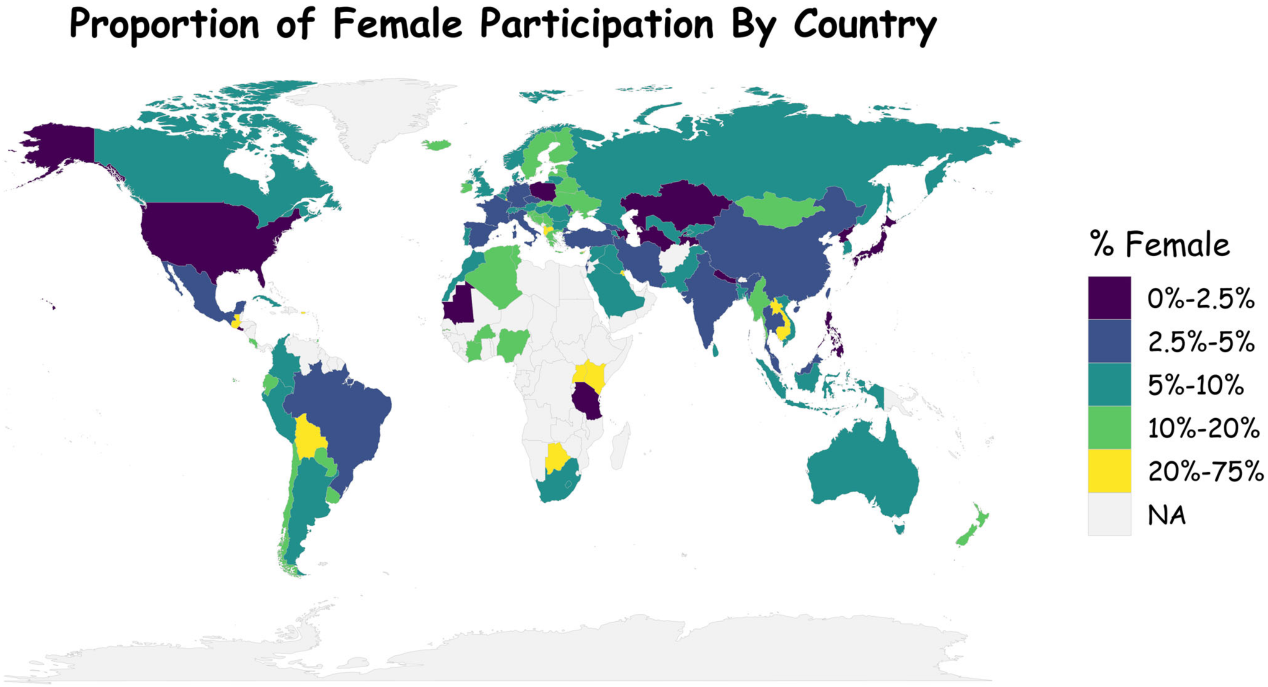

Figure 7. Countries are colored according to their overall proportion of female participants in the IMO.

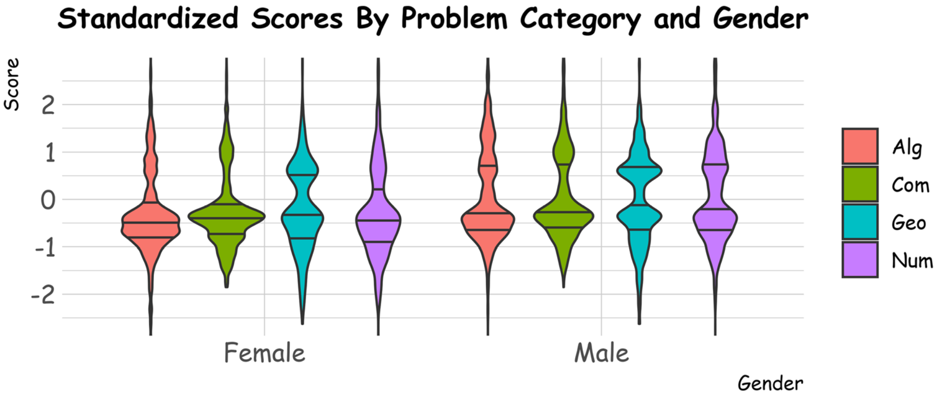

Figure 8. Standardized scores depicted in violin plots with 25%, 50%, and 75% quantiles marked by horizontal lines and stratified by problem category and gender. Female participants do much better on geometry problems relative to other problem types.

Here is the regression equation that was fit to the data with an adjusted R^2 of 0.26:

$$\begin{aligned}\text{Score} \sim \text{Gender} * \text{Problem.Type} + \text{Country} * \text{Year} \\ + \text{Country} * \text{Year}^2.\end{aligned} \tag{2}$$

Note that total contest scores and scores averaged over all players on a team will have less random variation than individual question scores, which explains why this adjusted R^2 value is somewhat lower than in the other model fits.

Another model like (2) was fit but with just additive terms for gender and problem type without an interaction term. Comparing the two model fits by analysis of variance showed that the addition of the interaction term made the model fit substantially more strongly, with a p-value of 7.7×10^{-6}. As seen in Figure 8, the interaction is dominated by the geometry problem type—the female participants do much better on these problem types, whereas the male participants also do slightly better but nowhere near to the degree of the female participants.

As a final comment on gender, we take a look at the "Hall of Fame" as published on the website imo-official.org, which is a ranking of IMO participants by medal count. Although only three female students are among the top 100 on this list, one of those students, Lisa Sauermann, sits quite near the top, at number 3, having the impressive record of scoring four gold medals and one silver medal for Germany, including one year with a perfect score. This is certainly an inspirational example to any aspiring mathematics student.

Discussion

In this analysis, we identified some interesting statistics and patterns from IMO contest data. The 2017 problems had both the easiest and hardest IMO problems of all time, and there seems to be a general trend in recent years that the easy problems are a little easier compared to years past and the hard problems are a little harder. We also found that problems 1 and 4, which are generally the easiest problems on the test, are most likely to be geometry problems. We also showed that countries demonstrated differential performance by problem type, with Greece's strength in geometry representing one of the most extreme examples. The USA early on was weaker in geometry compared to the other problem types, but in recent years geometry has been their forte. China, Russia, and the USA have the strongest teams overall, with China the historic powerhouse but with the USA edging out in front in the past few years. There seems to be a home advantage; that is, hosting the IMO is associated with a slight advantage by improving the host country's team score by approximately 9.4 points on average. Similarly, a problem submitted by one's own country is associated with slightly better performance on that problem. Finally, an analysis of gender differences shows that females are underrepresented in this competition, and they tend to do relatively better on geometry problems compared to other problem types.

Putting together the IMO contest takes a tremendous amount of effort. A typical year might have 150 problem proposals from 50 or so different countries that are then pared down to about 30 shortlisted problems by the host-appointed Problem Selection Committee. Then the problems are selected by majority vote, where each participating country gets a single vote. The final contest is then translated into nearly 60 different languages.[5] Although this paper identified various biases in the contest, the process of preparing the contest as a whole seems to be practical and quite fair.

However, one place for clear improvement in the contest would be to promote greater participation by females. One possibility would be to hold a female-only IMO contest alongside the IMO event. Examples of current all-female math contests include the European Girls Mathematical Olympiad, the China Girls Mathematical Olympiad, and the Math Prize for Girls. There is plenty of precedence for this kind of structure in chess: the Women's Chess Olympiad has been part of the Chess Olympiad events since 1976. A prestigious International Girls Math Olympiad would certainly greatly promote female participation in math problem-solving.

ACKNOWLEDGMENTS

The author appreciates the careful and detailed comments of the anonymous reviewers that led to a much improved manuscript. The author appreciates the efforts of Dr. Huamei Dong, who helped with classifying the problems, and Angelina Berg, who provided editorial support.

Arthur Berg
Department of Biostatistics and Statistics
Penn State University
Hershey, PA
USA
e-mail: berg@psu.edu

---

[5] More details on the IMO general regulations are available at https://www.imo-official.org/documents/RegulationsIMO.pdf.