Analysis of the results of the participants of the Republican Olympiad in Mathematics

---

Author: Adilet Malikov, Aziret Masalbekov

Date: 01.12.2023

Instructor: Polina Dolmatova

Abstract

This study presents an analysis of the results of participants in the Republican Olympiad in Mathematics (ROM) for the years 2020 and 2023, spanning grades 6 through 11. We aim to evaluate the effectiveness of participants in solving mathematical problems, considering variables such as type of problem, grade, year, and gender. The analysis involves hypotheses related to task complexity, participants' success rates over time, gender distribution, and correlations between gender and success rates. We utilize data sorting techniques, including classifying tasks by topic and sorting data by gender, to derive meaningful insights

Introduction

The main focus of our work was the results and tasks demonstrated during ROM-2020 and ROM-2023. The analysis shows how effectively the participants of the Olympiad are able to solve the problems and in which areas of mathematics the participants need additional help.

Main idea

The main goal and idea of this analysis was to help the participants and teachers to find out the weaknesses in which there are gaps in the knowledge of schoolchildren, which will help to fill these gaps and as a result will raise the overall level of all Olympiad participants. The topic of correlations of various aspects unrelated to the complexity of the Olympiad and the success of the participants is also considered.

Hypotheses and assumptions

H1: Task complexity is the same; differences in difficulty are minimal.

H2: The overall success rate of participants increases with continued training.

H3: The number of male and female participants is approximately equal.

H4: There is no or minimal correlation between gender and participants' success rate.

H5: The solve rate of multiple-choice and free response problems is the same or the differences are minimal.

H6: The average percentage of successful problem solving by Olympiad participants increases over time.

Initial data

Initially, we had data collected during the Olympiads from 2020, 2021, and 2023, for grades 6 through 11. Not all years contain complete information and individual aspects vary from year to year. We considered the data collected for 2020 to be the most informative for the purposes of our analysis, due to the availability of not only the number of correct answers for a single participant but also the availability of their answers to each task. The data for 2023 contains the number of correct answers, but unfortunately, we do not have access to the answers of participants in sixth, seventh, and eighth grades, so we are unable to determine which section of mathematics is used in their solution, but we do have data for ninth, tenth, and eleventh grades. The data for 2021 contain only the number of correct answers of the participants but not their answers and we also do not have access to the tasks.

We analyze the data of all participants of the Olympiad. Analysis of all possible data allows us to get the most reliable result of the analysis and make correct conclusions and advice.

Data sorting

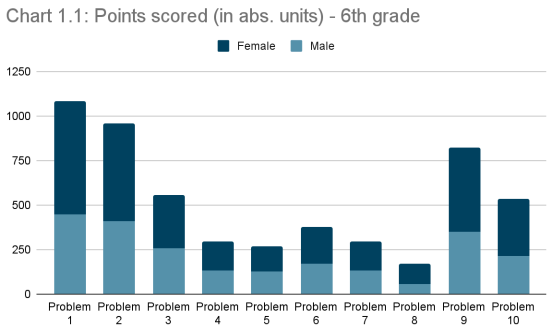
The data are presented in the form of tables containing several rows with different information about the Olympiad. The group of tables was determined by the year of the Olympiad and by the year of training of the participants. Finding correct answers was carried out by the participant or participants who scored the highest possible score at the Olympiad. The answers of the participants who scored the maximum possible score at the Olympiad were considered correct for each task. Taking into account the correct answer, we counted the number of other participants with the identical answer and this number is the total number of pupils who successfully coped with the task. We repeated this process among all possible

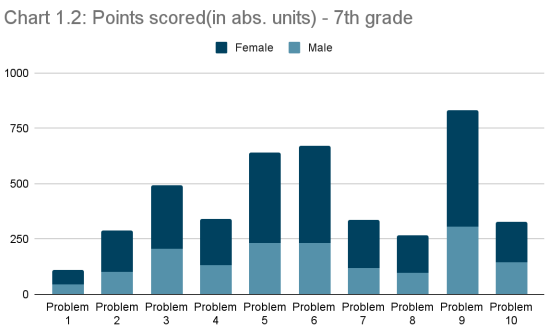
grades and years given the original data. In order to determine what topic a task from the Olympiad uses for further analysis, we used the tasks that were given to the participants, then we analyzed the tasks and assigned a topic for each one, and then assigned the number of people who successfully completed it. In this way, we managed to get the topics of the Olympiad tasks and how many pupils were able to solve them correctly.

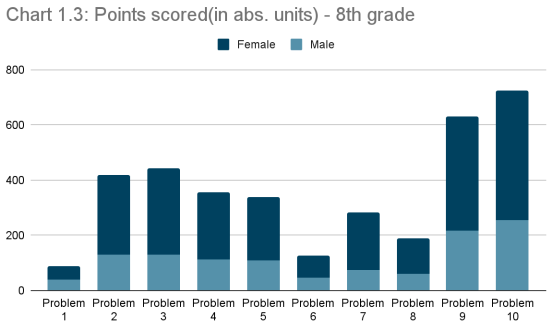
Sorting by gender involved only counting the number of male or female designations in each individual class. The value of the participant's gender was already included in the raw data. All we had to do was to use Excel to calculate the number.

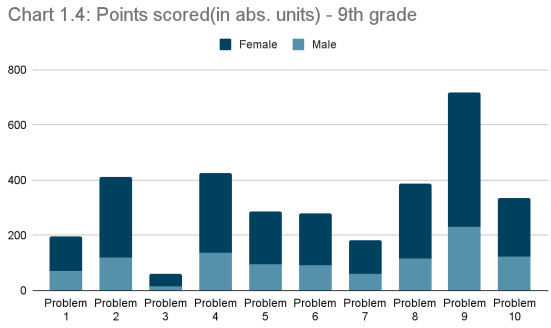
Results

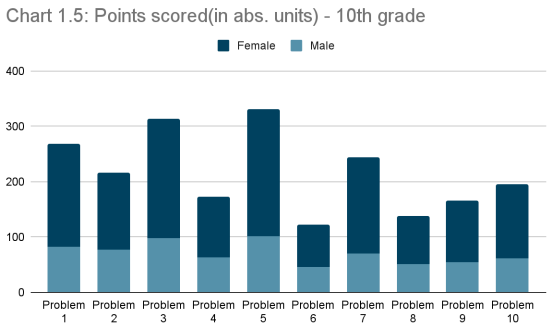
These graphs show the results of the participants in all classes. The number of solvers is shown in absolute values for each individual problem. The different colors of each column divide the solvers into males and females.

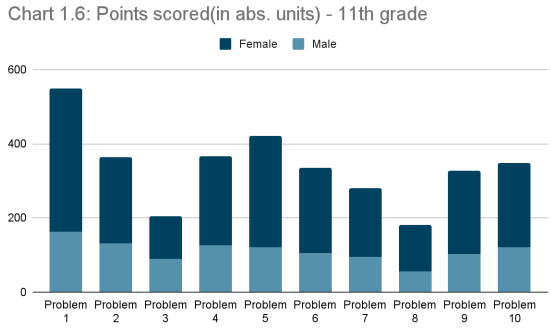






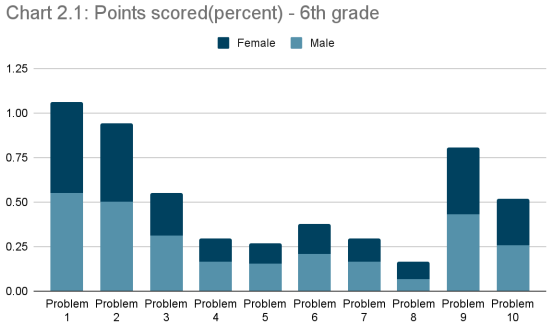


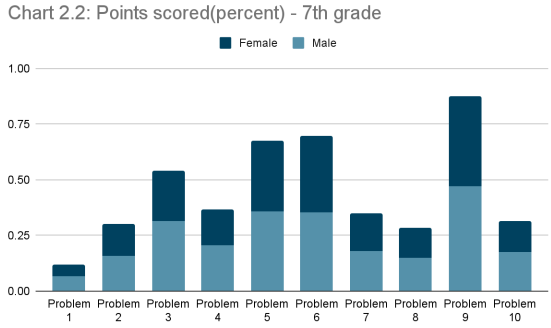


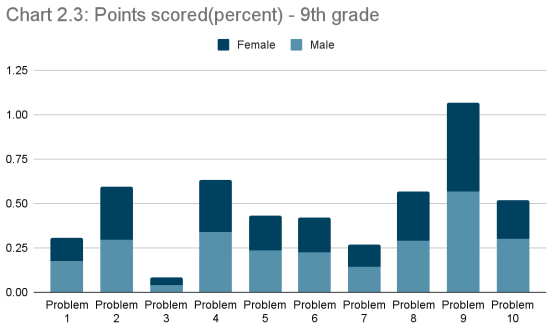


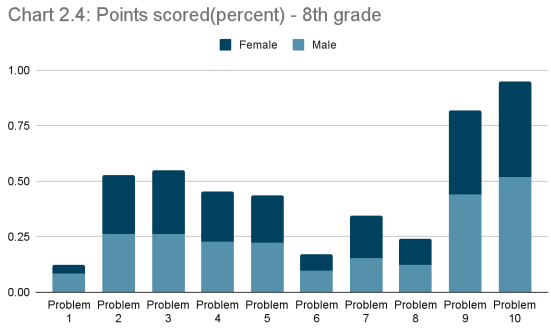
Considering charts 1.X: we can conclude that the number of problem solvers varies by gender and in absolute values, the number of women exceeds the number of men

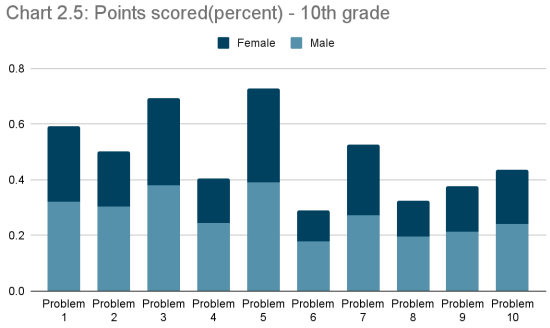
And for data from 2023 we can see the amount of participants who solved problems in absolute units.

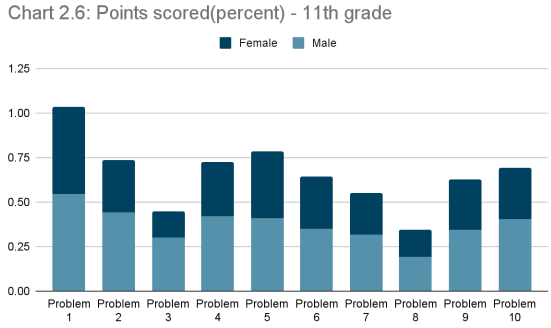






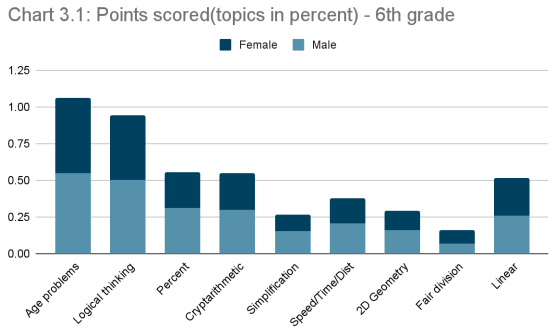


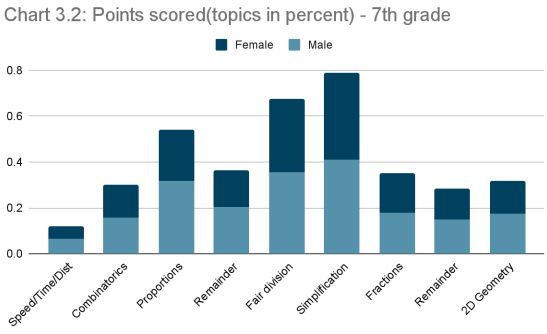


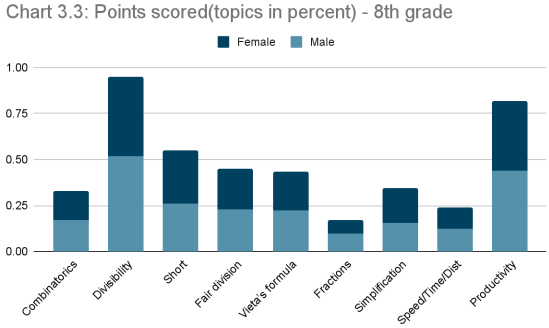


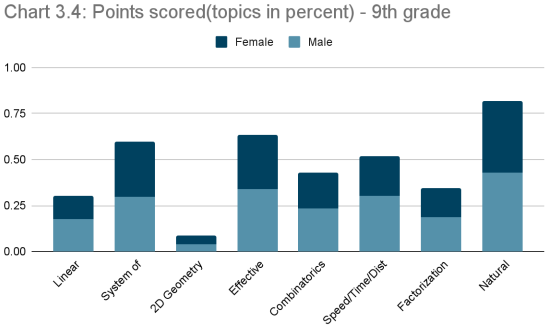
Considering charts 2.X: we can conclude that although the number of women solving problems is higher, the percentages show that men have a slightly higher percentage of successful solutions than women.

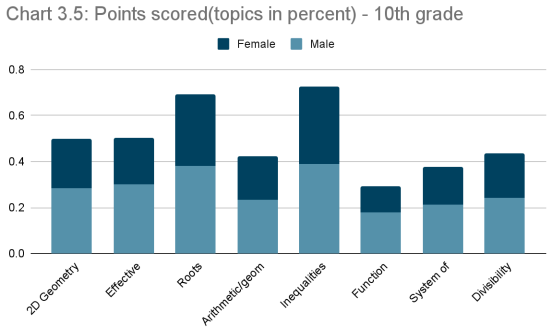
And for data from 2023 we can see the percentage of solvability of problems:

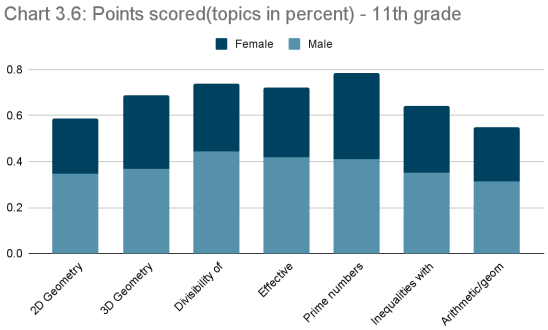












Considering charts 3.X graphs: we can see that the topics differ according to class, males have a higher percentage of solutions than females. Further conclusions have to be made individually for each chart due to the different topics.

Chart 3.1: The chart clearly indicates the problem areas of the participants in the tasks: Fair Division, 2D Geometry, and Simplification. While in the tasks with Age Problems and Logical Thinking, the participants perform well.

Chart 3.2: The chart clearly indicates the problem areas of the participants in the tasks: Speed/Time/Dist, Remainder, 2D Geometry. The problem areas are similar to the previous year of study in the 2D Geometry and Speed/Time/Dist topics. While in tasks with Fair Division and Simplification, participants perform well.

Chart 3.3: The chart clearly indicates the problem areas of the participants in the tasks: 2D Geometry, Fractions, and Combinatorics. Whereas in Divisibility and Productivity tasks, participants perform well.

Chart 3.4: The graph clearly indicates the problem areas of the participants in the tasks: Linear equations and 2D Geometry. The problem areas are similar to Chart 3.1 and Chart 3.2 of the training in the 2D Geometry topic. In the Natural numbers and Effective calculation of large numbers tasks, participants performed well.

Chart 3.5: The graph indicates the problem areas of the participants in the tasks: Function. This one does not have a strong difference in the level of successful solutions. However, the tasks with the themes of Roots and Inequalities stand out for their high percentage of successful solutions.

Chart 3.6: This chart does not have a sharp contrast in the success rate of tasks. Each topic has a relatively high solution rate. This leads to the fact that the general level of almost all participants allows them to solve all tasks.

Chart 3.7: The graph indicates problem areas in Math: Symmetry and Circles. And catastrophically low percentage of solving problems on Function composition. There are high solution rates for Consecutive natural numbers and System of trigonometric equations.

Chart 3.8: The chart indicates problem areas in the sections of mathematics: Triangles, System of equations, Effective calculation of large numbers and Stereometry. And catastrophically low percentage of solving problems on Function composition. Similar problems to Chart 3.7 grade can be seen in Stereometry and Function composition. There are high levels of solutions to problems on the topics Optimization and Cryptarithmetic. And there is a distinctively high percentage of solutions for Speed/Time/Distance.

Chart 3.9: The chart indicates problem areas in the sections of mathematics: Diophantine problem, Effective calculation of large numbers, Percents. And catastrophically low percentages of solving problems on Function composition and

Triangles. Similar problems to Chart 3.8 grade can be seen in Effective calculation of large numbers, Triangles and Function composition. High levels of solutions to problems on the topics Trigonometric inequality and System of equations are observed. And a distinctively high percentage of solutions to problems on the topic Cryptarithmetic.

Chart 4.1: Number of male and female participants

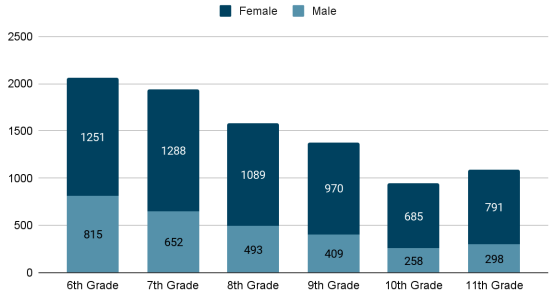


Chart 4.2: Problems solved in average by grades and gender

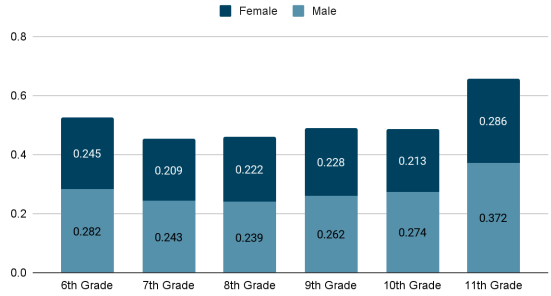


Chart 4.3: Problems solved in average

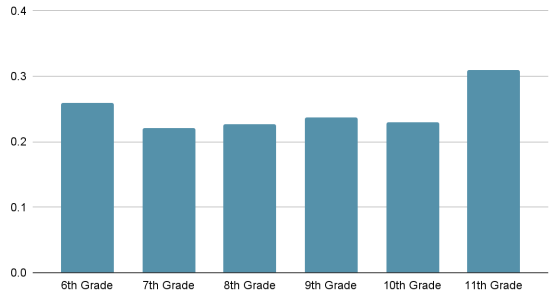


Chart 4.X: These are charts showing data not for a particular year of study but data from all years and taking into account the number of participants.

Chart 4.1: According to this chart we can conclude that there are more female participants than male participants in all the examples.

Chart 4.2: This chart indicates that despite the superior number of male participants, men solve more problems at the Olympiad in percentage terms.

Chart 4.3: The graph shows the success rate of problem-solving in each of the years of study, with the highest results in classes number six and number eleven.

Discussion and conclusion

H1: Task complexity is the same; differences in difficulty are minimal.

Taking into account charts 1.X and 2.X, we can conclude that the success rate of participants is completely different and in various problems we can see extreme change of success rate which lead us to conclude that task complexity is also very different and that hypothesis is false.

H2: The overall success rate of participants increases with continued education.

In order to confirm or refute this hypothesis we should pay attention to chart 4.3 which indicates the success rate of problem solving in the progression of years of education. We can notice that the grades from seventh to ninth do show a tendency to increase the success rate of solutions, but the transition from sixth to seventh on the contrary indicates a decline, as well as the transition from ninth to tenth. After that in the final eleventh grade, there is a clear jump in the growth of successful solutions. It follows that despite the fact of an increase in the success rate, we cannot say that there is a linear improvement with the progression of learning.

H3: The number of male and female participants is approximately equal.

To confirm or refute this hypothesis we should look at chart 4.1 the graph clearly shows that the number of females exceeds the number of males in each of the years of study. Despite the decrease in the total number of participants, the trend in the difference between male and female participants remains the same.

H4: There is no or minimal correlation between gender and participants' success rate.

To confirm or refute this hypothesis, we should look at chart 1.X, 2.X, 3.X and 4.2.

If we pay attention to charts 1.X: It can be seen that the number of women who solved the tasks exceeds the number of men, from which we can conclude that the chance of successful solution of the task by a woman is higher, but when we consider charts 2.X: These charts contain the success rate not in absolute values but as a percentage of the total number of participants of a certain gender. It can be seen that the percentage of successful problem solving by men is little higher than by women. This conclusion is confirmed by charts 3.X: These graphs contain the percentage of successful solutions of problems by sections of mathematics and indicate the higher number of successful solutions on the part of men.

Also, Chart 4.2: The graph indicates the average percentage of problem solving by gender for each year of study. The graph confirms that there is a different percentage of successful solutions depending on gender.

Taking into account all the data we can conclude that despite the overall superiority of women solved problems in absolute values compared to men, when considering the total number of participants of each gender and percentage consideration we can conclude that men have a higher percentage of successfully solved problems, which leads to the refutation of the hypothesis that there is no correlation between gender and the percentage of successful problem solving.

H5: The solve rate of multiple-choice and free response problems is the same or the differences are minimal.

To confirm or refute this hypothesis we can look at the data that Charts 1.7, 1.8, 1.9 and 2.7, 2.8, 2.9 provide. These charts indicate a catastrophic drop in the percentage of problem solving if the problem requires free response rather than multiple-choice answer. Also Graph 5.1 indicates a huge difference between

multiple choice and free response question types, a trend that continues across all grades.

H6: The average percentage of successful problem solving by Olympiad participants increases over time.

To examine the validity of this hypothesis, we can refer to Chart 6.1, which clearly shows the average level of ROM-2020 and ROM-2023, as follows 2023 showed a strong decrease in the average problem solving rate for all classes. We believe that this decline is partly due to the introduction of a new type of problems that require free response answers, which significantly reduces the solvability rate as follows from *H5*. Therefore, we decided to look at the level and differences between the different years without the innovation, the analysis of ROM-2023 without free response answers is shown in Chart 6.2, which shows that even in the absence of the free response answers component, the success rate of ROM-2023 is lower than ROM-2020 and we can conclude that the success rate has not increased over time and therefore the hypothesis is false.

One of the possible reasons for decrease of success rate over 2020-2023 years is global pandemic and its influence on the quality of education. The disruption caused by pandemic, such as shift to remote learning, may have influenced the participants' preparation and overall performance in the Olympiad.

Predictions

Sustained Gender Difference

We anticipate that, despite the higher number of female participants, men will continue to have a slightly higher percentage of successful solutions, as observed in Chart 4.2.

Predict that efforts to bridge the gender gap in certain sections may yield some improvements but may not completely equalize success rates between genders.

Prediction 2: Fluctuating Success Rates with Grade Progression

Consistent Problem Areas

Anticipate that the identified problem areas in specific topics, such as Fair Division, 2D Geometry, and Simplification (Chart 3.X), will persist in future Olympiads.

Predict that continued attention to these problem areas may lead to incremental improvements over time.

Unequal Distribution of Male and Female Participants

Predict a continued trend of more female participants than male participants in all grade levels, as shown in Chart 4.1.

Anticipate that despite efforts to encourage participation from both genders, the unequal distribution may persist.

Non-linear Improvement with Continued Education

Expect a non-linear improvement in the overall success rate of participants with continued education, as observed in Chart 4.3.

Predict that certain grades, such as the sixth and eleventh, will consistently show higher success rates compared to others.

These predictions are based on the trends and patterns identified in the data analysis and serve as expectations for the future performance of participants in the ROM.

Impact of Free Response Questions

Expect that the inclusion of free response questions in the Olympiad may continue to affect the overall success rate, as observed in Chart 6.1. Predict that participants may face challenges in adapting to this format, and organizers may need to consider refining the balance between multiple-choice and free response questions to maintain the integrity of the assessment.

Recommendations

Our recommendations are based on data collected from chart 3.X:
There are several ways to solve the problem areas of the participants.

Increase the amount of time spent on problematic areas, we have already identified problematic areas in math and based on our results we should spend more time on them, more information on problem areas can be found in the Results section and by searching for a specific year of study. Identify specific math topics and problem areas that consistently show lower success rates among participants as shown in chart 3.X. Allocate additional instructional time and resources to address these problem areas. Develop targeted instructional materials or workshops to increase understanding and knowledge of these specific topics.

Another possible way is to determine different amounts of points for certain sections of the Olympiad, for example, increasing the points for problem areas due to their complexity, which would also give an additional incentive for participants to pay more attention to the problem areas. Consider adjusting the scoring system to reflect the difficulty of different sections or topics of the Olympiad. Give higher scores to tasks in areas where participants typically struggle to encourage more attention and effort in those specific topics.

Recognize gender differences in success rates as shown in Tables 1.X, 2.X, and 3.X. Develop gender-specific instruction and training strategies for participants to account for observed differences in male and female performance. Implement initiatives to encourage female participants and close the gender gap in some sections of the Olympiad.

References:

Республиканские математические олимпиады 2020-21 и 2021-22 гг. Задачи и решения: Учеб. Пособие. - Б.:2022. - 122 с., Авторский коллектив, 2022

LibreOffice Calc Functions: https://help.libreoffice.org/6.1/en-US/text/scalc/01/04060100.html

R for Beginners, 2002, 2005, Emmanuel Paradis (12th September 2005), https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

What is Data Analysis?: Process, Types, Methods, and Techniques: https://www.simplilearn.com/data-analysis-methods-process-types-article

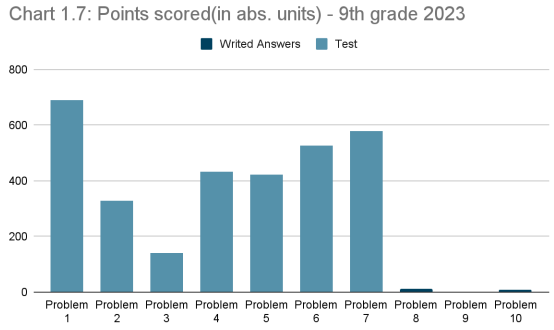
Descriptive Statistics: Definition, Overview, Types, Example: https://www.investopedia.com/terms/d/descriptive_statistics.asp

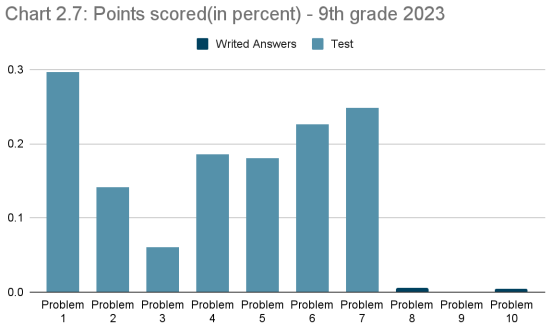
KSMS (Кыргызско-Шведская математическая школа): https://ksms.kg/instrukcia-dlya-uchastnikov-rom

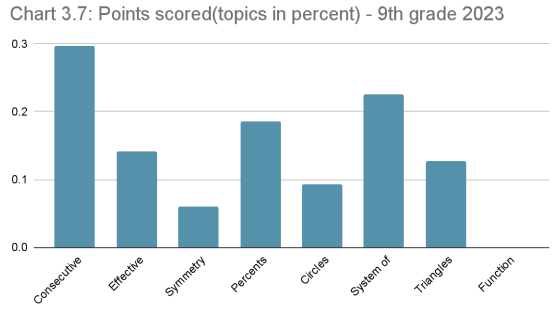
GitHub with all the data: https://github.com/adiletmalikov/prob-project

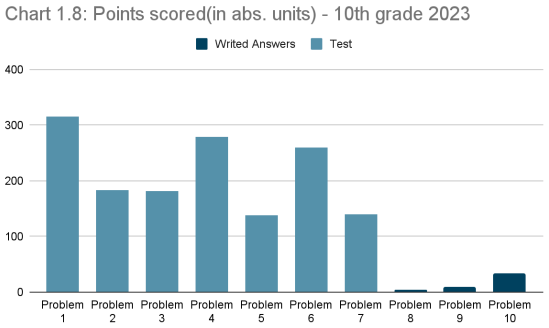
ADDITIONAL DATA

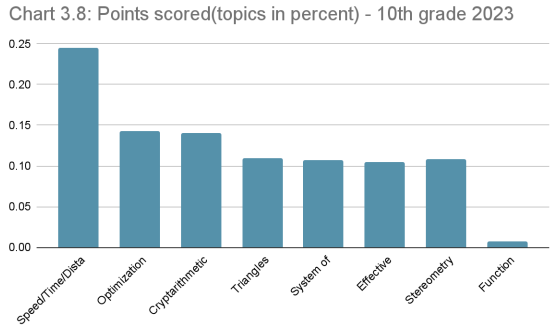
CHARTS:











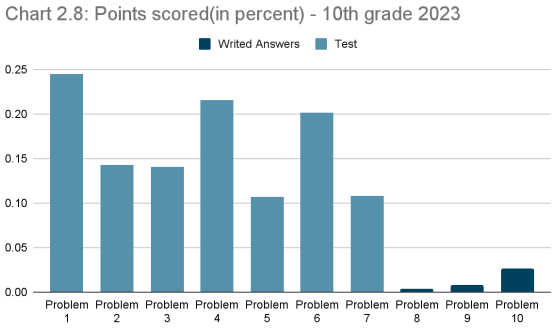


Chart 1.9: Points scored(in abs. units) - 11th grade 2023

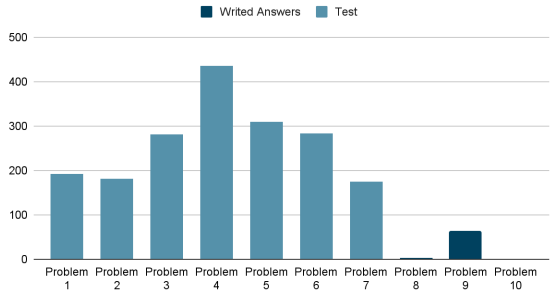


Chart 2.9: Points scored(in percent) - 11th grade 2023

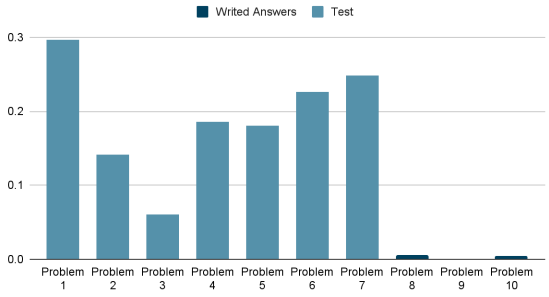


Chart 3.9: Points scored(topics in percent) - 11th grade 2023

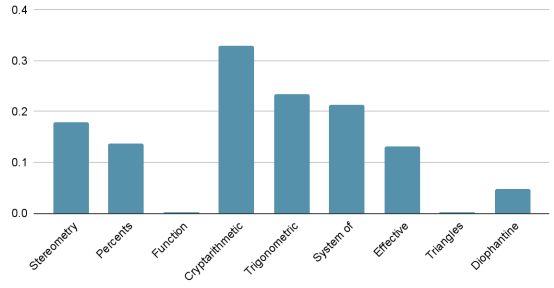


Chart 5.1: Solve rates of multiple-choice and free-response problems

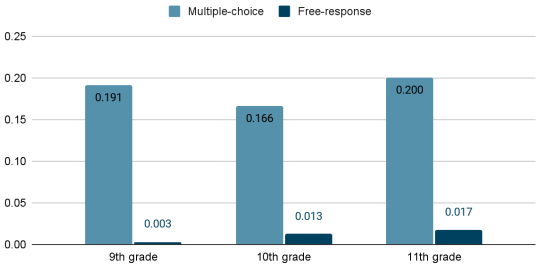


Chart 6.1: Solve rates by 2020 and 2023 years (of all problems)

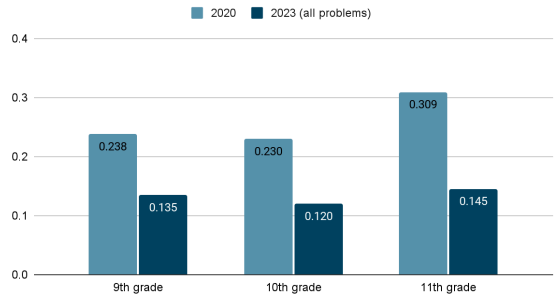
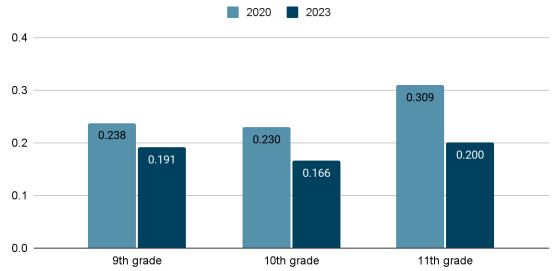


Chart 6.2: Solve rates by 2020 and 2023 years (of multiple-response problems only)



EXCEL TABLES:

2020

6th grade:6th grade 2020.xlsx

7th grade:7th grade 2020.xlsx

8th grade:8th grade 2020.xlsx

9th grade:9th grade 2020.xlsx

10th grade:10th grade 2020.xlsx

11th grade:11th grade 2020.xlsx

2023

9th grade:9th grade 2023.xlsx

10th grade:10th grade 2023.xlsx
11th grade:11th grade 2023.xlsx