STUDENT PERFORMANCE – TEST SCORES ANALYSIS

LAKSHMI MUTHUKUMAR

Overview of the problem

- The performance of high school students from all over the U.S. is under analysis
- The data comes from Kaggle with 1000 observations and 8 variables.
- How does their scores differ with respect to difference in the data variables on an average?
- For example, how male and female students fare in these different categories? Is there a significant difference in their relative performance? If so, how did they differ on an average?
- Visual analysis, descriptive statistics, Shapiro-Wilk, independent sample t-test for hypothesis testing were done

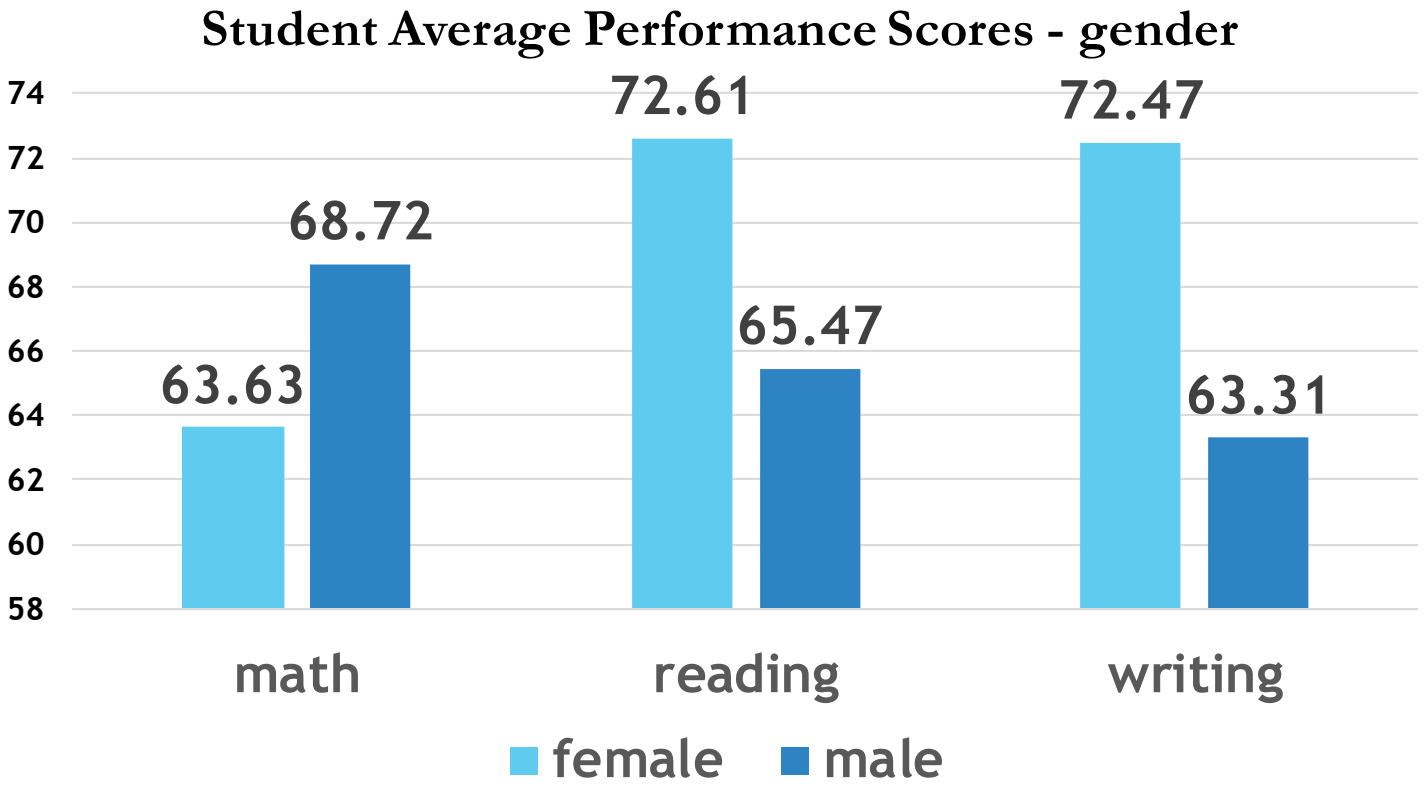
Overview of the data

- **Variables studied:** gender, race/ethnicity, parental level of education, lunch mode differences, and test preparation course
- Scores of the students in the following categories
    - A. Reading
    - B. Writing
    - C. Math
- **Gender:** Students in this dataset are two genders namely, **male and female**. How do students fare in the above three categories with respect to gender?
- **Race/ethnicity:** Students in this dataset are spread across various race/ethnicities. How do students fare in the above three categories with respect to it? There are five different racial groups namely, **group A, group B, group C, group D, and group E.**
- **Parental level of education:** Students in this dataset are spread across with parents having different levels of education. How do students fare in the above three categories with respect to it? There are six different levels of education in parent's of the students namely, **high school, some high school, some college, associate's degree, bachelor's degree and master's degree.**
- **Lunch Mode Differences:** Students in this dataset are spread across different lunch mode differences. How do students fare in the above three categories with respect to it? Students can either opt for **Free/Reduced** lunch option or **Standard option**.
- **Test Preparation Score:** Students in this dataset are spread across different test preparation results. How do students fare in the above three categories with respect to it? Students either **Completed or not(None)** test preparation courses.

HYPOTHESES

- **a. Gender:** *Ho:* There is no significant difference between the Reading, Writing, and Math scores of male and female students. *Ha:* There is a significant difference between the Reading, Writing, and Math scores of male and female students.
- **b. Race/Ethnicity:** *Ho:* There is no significant difference between the Reading, Writing, and Math scores of students belonging to various race and ethnicities. *Ha:* There is significant difference between the Reading, Writing, and Math scores of students belonging to various race and ethnicities.
- **c. Parental Level of Education:** *Ho:* There is no significant difference between the Reading, Writing, and Math scores of students with parents with different levels of education. *Ha:* There is significant difference between the Reading, Writing, and Math scores of students with parents with different levels of education.
- **d. Lunch preferences** *Ho:* There is no significant difference between the Reading, Writing, and Math scores of students with lunch option preference. *Ha:* There is significant difference between the Reading, Writing, and Math scores of students with lunch option preference.
- **e. Test Preparation Course** *Ho:* There is no significant difference between the Reading, Writing, and Math scores of students with test preparation course variable difference. *Ha:* There is significant difference between the Reading, Writing, and Math scores of students with test preparation course variable difference.

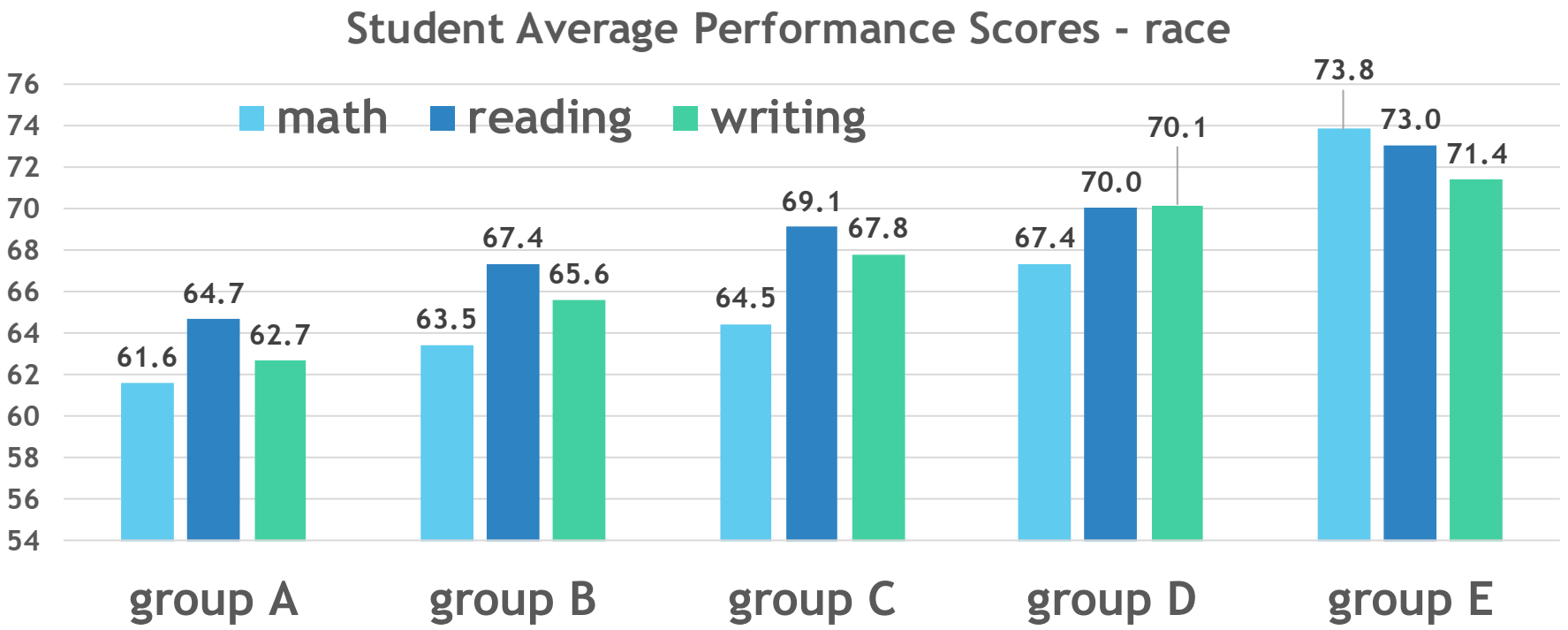
Gender



There is significant difference in the scores w.r.t. gender

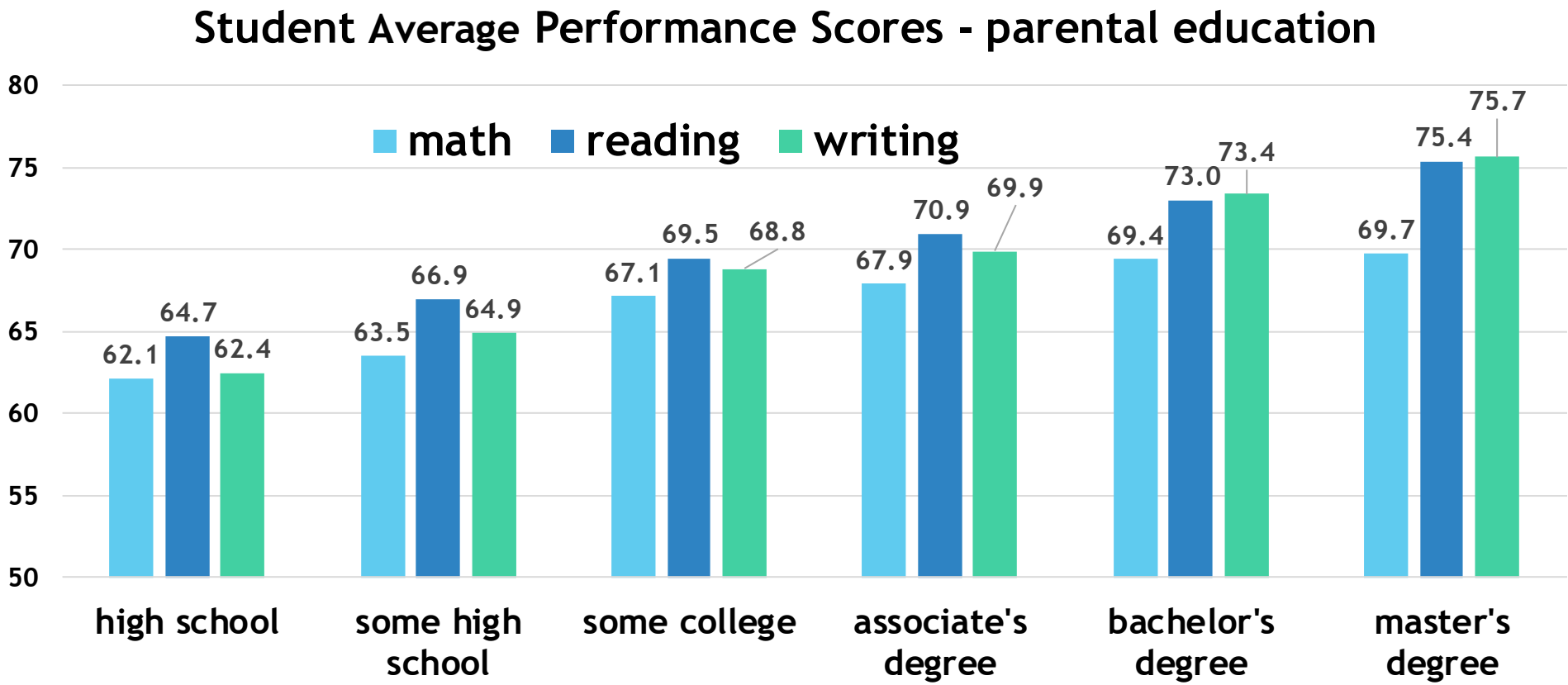
- In **Math**, male students are performing better than female students
- In **Reading**, female students are performing better than male students
- In **Writing**, female students are performing better than male students
- The **average difference** for math, reading and writing are **5.1, 7.2 and 9.1** respectively
- **Inference: Women do better in literature and men do better in engineering/allied**

Race



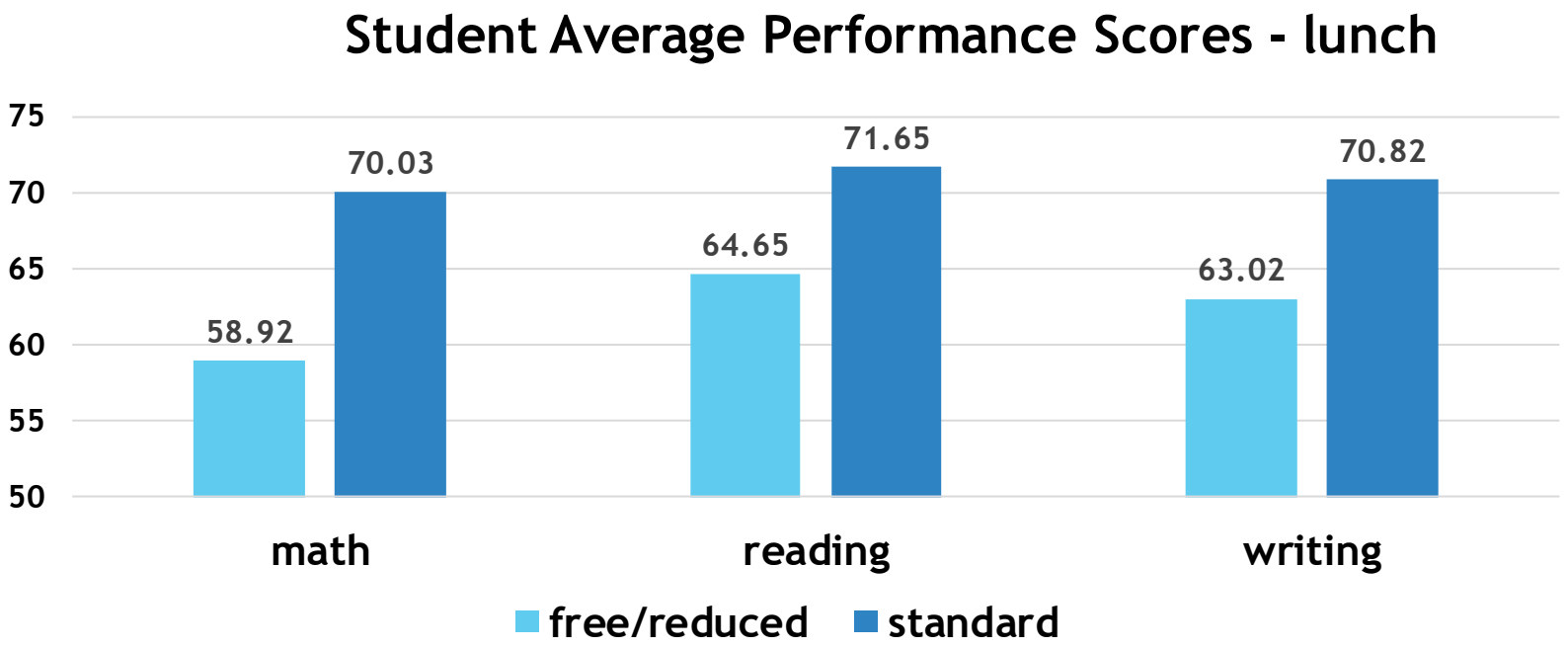
- **Group E outperforms** every other group in all three scores. Performance of **group A is the lowest**.
- In both **Math and Reading scores**, group **E is different** from other groups.
- In **Writing scores**, **A, and B are similar** and **D, E are similar** to each other.
- The **average difference** in highest and lowest scores in math, reading and writing categories are **12.2, 8.3 and 8.7.**
- **Inference: All races perform differently in all the different categories.**

Parental level of education



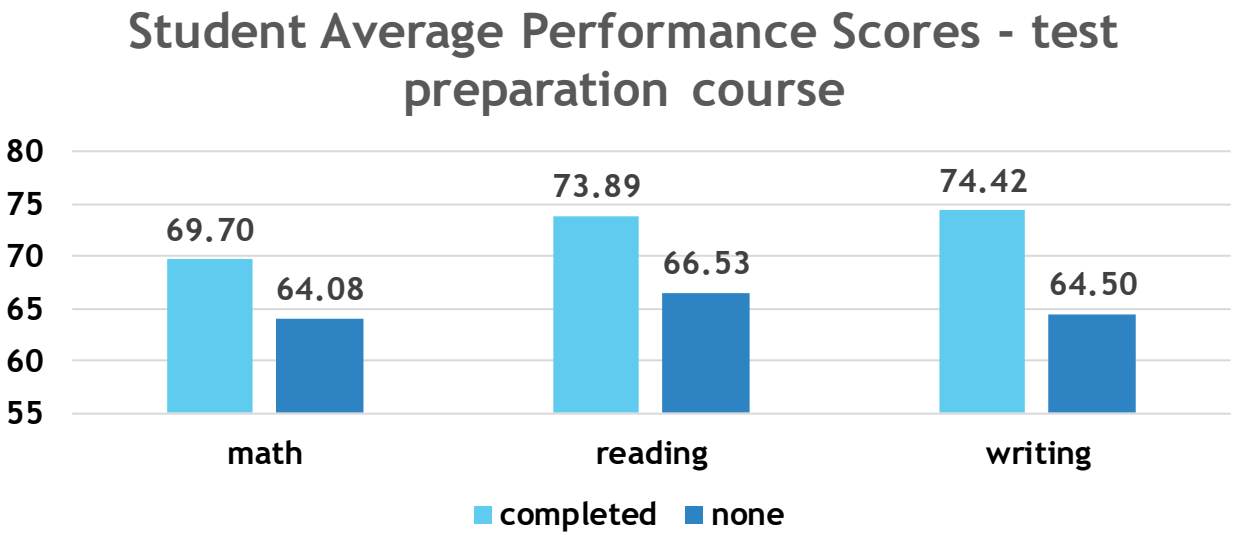
- Students with parents having **Masters degree outperform,** with that of **high school** is the **lowest**
- In **Math**, students with parents having any **college degree is different** from those with only high school degrees
- In **Reading**, students with parents having **Masters, Bachelors, or Associate degree is different**
- In **Writing**, with parents **Bachelors and Masters degree** is different from those with only high school degrees
- The **average difference** in highest and lowest scores in for math, reading and writing are **7.6, 10.7 and 13.3**
- **Inference: Parents with higher degree of education have children with better performance in all categories.**

Lunch Preferences



- There is significant difference in the scores of students with Lunch preference in all the three testing categories.
- The students with Lunch preference **"Standard"**, have **better scores** in all three categories
- The **average differences** are approximately found to be **11, 7 and 7.8** respectively for the lunch variable in math, reading and writing categories
- **Inference: We can infer that students with higher household incomes have better performance in the categories**

Test Score preparation



- The two groups with Test preparation course complete and none, have significantly different scores.
- The **completed group performs better** in all categories of the test.
- The **average differences** are approximately found to be **4.9, 7.3 and 10** respectively for the test preparation course variable in math, reading and writing categories
- **Inference: We can infer that students with higher motivation to complete the test preparation course have better performance in the categories**

Recommendations

- Women do better in literature and men do better in engineering/allied
- All races perform differently in all the different categories
- Parents with higher degree of education have children with better performance in all categories
- We can infer that students with higher household incomes have better performance in the categories
- We can infer that students with higher motivation to complete the test preparation course have better performance in the categories

Conclusions

- Sampling bias is common in such studies
- Sample size for such a geographically diverse group should be considered
- Data is generic, a location specific data might help improve the analysis
- In addition to the current variable, internet access or access to external help can be an important variable

THANK YOU!