MIDDLE SCHOOL GRADES PROJECT

ALY 6070 Communicate/Visual Data Analysis

PRERAK SHAH, CHIRAG SHAH

Northeastern University



Note:

Prerak Shah, Chirag Shah, College of Professional Studies, Northeastern University, Boston, MA 02115.

This report was created as a part of the project to gain hands-on experience in probability distribution function using R functionality.

Prerak Shah, Chirag Shah is now a student at Department of Analytics, Northeastern University

Contact: [shah.chir@husky.neu.edu](mailto:shah.chir@husky.neu.edu), shah.prer@husky.neu.edu

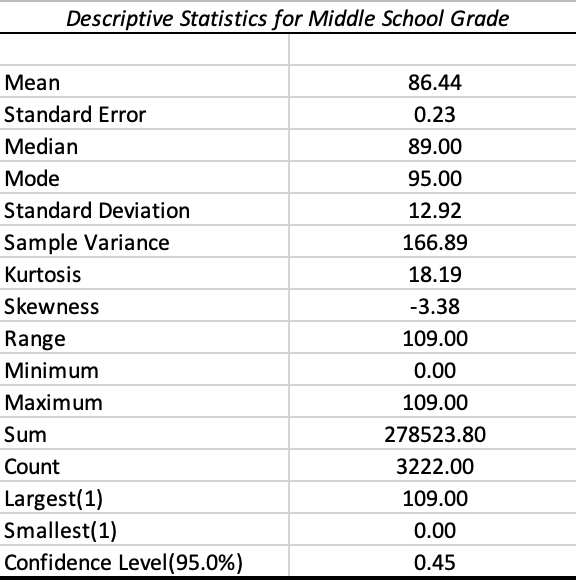
**INTRODUCTION**

We choose to analyze the dataset of “Middle School Grades” that was provided to us by our Professor. We chose this dataset from the others in specific, since we were interested in learning the middle school performances of the students. So, we can know that how good the students are from that school. There are 3222 grades given in the dataset. But we chose to analyze the sample of only 6th and 7th grade. We chose these sample size because we wanted to see that how the students with both statuses of ELL (English language learners) and SPED (Special Education) would particularly perform in their initial middle school. There are total of eight columns in which two columns namely performance letter grade and performance percentage grade are the dependent variables. We made efforts to analyze that the 7th grade students with ELL and SPED status score more than C+ letter grade than 6th grade students or not. The other question we tried to analyze is the students got less than 50% in 7th grade than in 6th grade. We have used excel as our statistical tool and Tableau as our visualization tool.

**ANALYSIS**

**PART-1: Descriptive Statistics**

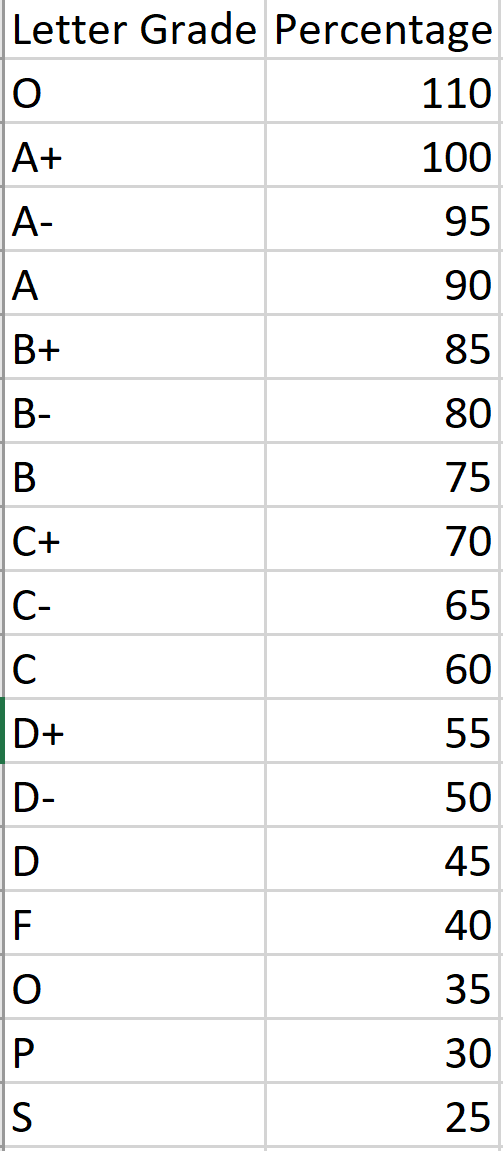
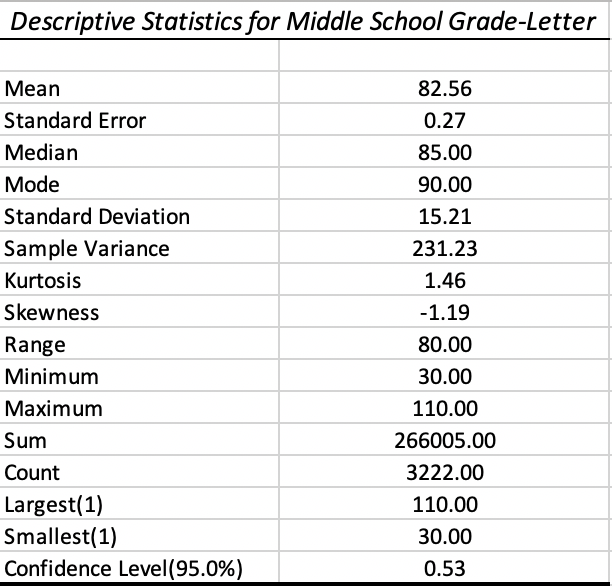
1. Descriptive statistics is shown below for the whole dataset. We chose to excel to develop statistical analysis and visual presentation as excel is convenient for data entry and for rapid row and column manipulation. As there is only two dependent numeric value column named Performance percentage grade, so we took out the exploratory statistics for only one column.



**Figure 1.0**- Descriptive Statistics for Middle School Grades

From the above table we can infer that the average percentage grade got by the students is 86.44%. The students who got 0%, might have not appeared to the exam. There are a greater number of students who got 95% in middle school.

1. From the below figure(left-side) we can observe that common letter grade is at least B-The most common letter grade got by the students is A-



**Figure 1.1**- Descriptive Statistics for Middle School Grades

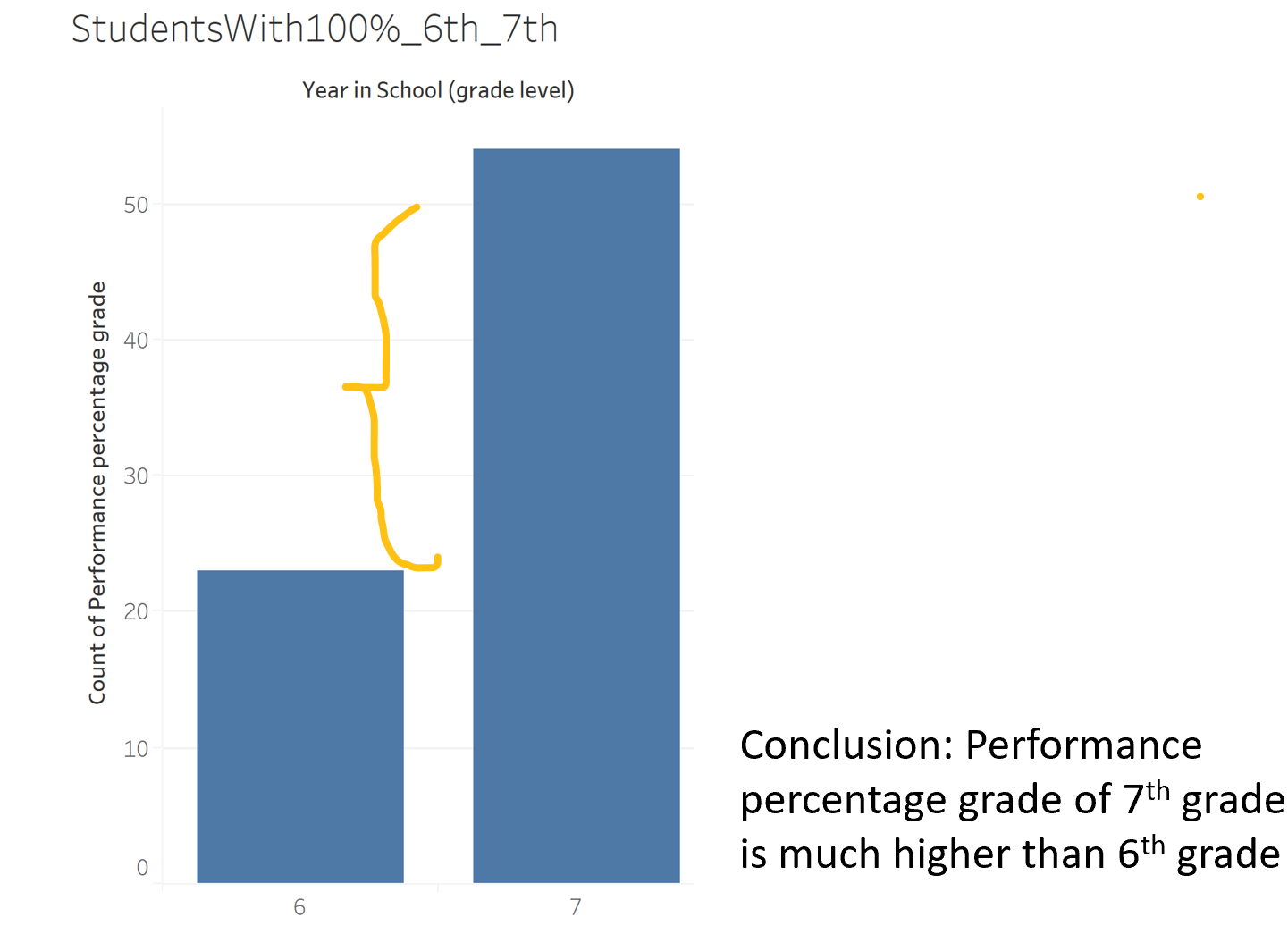
**PART-2: Inferential Statistics**

1. Students with Not\_SPED status in 6th grade and 7th grade is shown in the below figure. We can see that there is more dense area of performance percentage in grade 6.



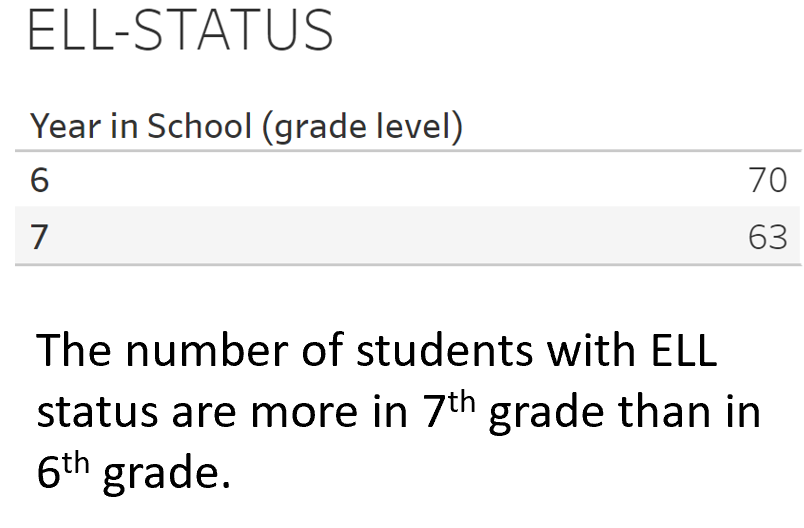
**Figure 1.2**- Scatter plot showing students with Not\_SPED status in 6th grade and 7th grade

1. Students with 100% grade in 6th grade vs 7th grade is shown below in the bar chart.



**Figure 1.3-**Bar graph showingstudents with 100% grade in 6th grade vs 7th grade

1. The figure showing the number of students with ELL status in 6th and 7th grade:





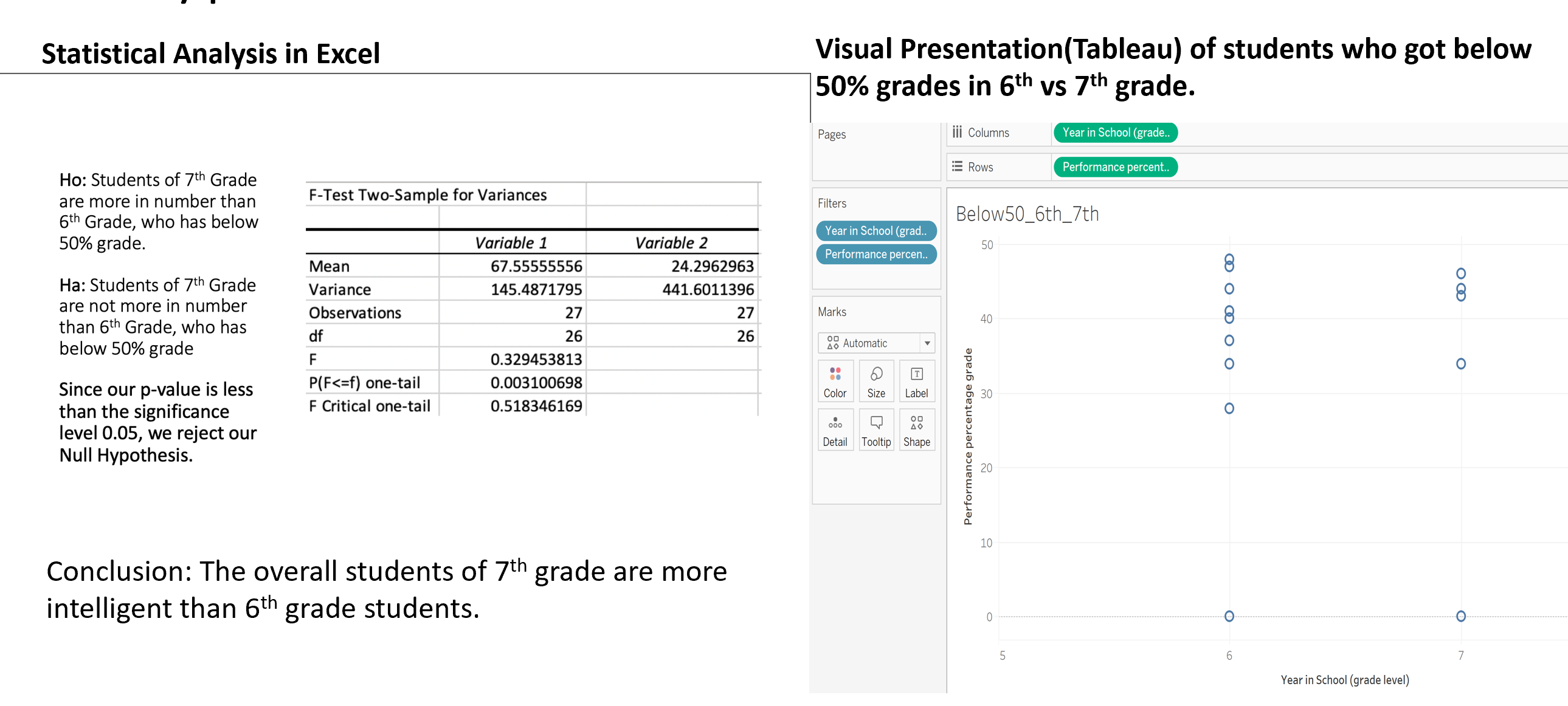
**Figure 1.4**- Table showing the number of students with ELL status in 6th and 7th grade



We found that there is a **positive correlation** in percentage grade vs grade letter for the middle school grade datasets. The correlation coefficient comes out to be **0.71**, which determines the coefficient of determination as 50%, which indicates the variability of response data i.e. percentage grade around the mean of the sample.

**HYPOTHESIS**

1. The below hypothesis is built in F-test:

**Figure 1.5**-Hypothesis with F-test

1. The below hypothesis is built in t-test:

**Figure 1.6-** Hypothesis with t-test

**CONCLUSION**

7th grade students performed excellent in the examination than the 6th grade, with a greater number of A+ and O grades. The current ELL and SPED status of students for both the grades have scored A grade in their examination. The coefficient of determination indicates that the variability in the performance grade with 50% response data around the mean.

**REFERENCES**

1. Wayne L Winston. (2013). Microsoft Excel 2013- Data Analysis and Business Modelling Ali Grami (April 2019) Probability, Random Variables and Random Processes. Page 2 - Introduction from Casino Freak.com
2. Korean J. Anesthsiol (2015). – T test as a parametric statistic. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4667138/>