Can Money Buy Brains?
A statistical analysis of Washington State teacher salaries and SAT scores

Ming Fong, Halle Hwang, Rae Gerking, Ethan Dang

January 24, 2020

1 Abstract

It's often said that *money can't buy happiness*. Yet in today's economic society, money seems to remain the driving factor in American lives, especially in the education system. While the United States remains the wealthiest nation in the world, it ranks 27th in healthcare and education, trailing behind a host of countries including Iceland, Finland, and the Netherlands. Furthermore, there has been a long debate in the American education system concerning teachers' salaries and whether or not they are paid enough. To this day, many educators feel that they have been denied competitive, professional pay for too long and now, with 2018 being the biggest year of teacher strikes in this generation, the effects could have an impact on the performance of American students. In this retrospective study, we will observe these issues by researching and analyzing average teacher salaries throughout public high schools in Washington state and directly compare them with average student test scores from the respective schools. Our goal in this study is to determine whether or not there is a direct association between teacher salaries and student performance.

2 Data Collection

We collected our data using online government datasets that documented average SAT scores[1] by school district for the 2018-2019 school year, as well as the salary of every public-school teacher[2]. There are 208 data points for the state of Washington where both a district's mean SAT score and secondary teacher salaries were reported. Possible sources of bias include the fact that we only have salary data for public schools, some school districts did not report mean SAT scores and some teacher salaries may have been unreported. In order to draw the most accurate conclusion, we used a large dataset to counteract these biases. The original datasets can be found in the appendix.

---

[1]SAT data: https://github.com/evilpegasus/APStats/blob/master/WA%20_District_SAT_2019.xlsx

[2]Salary data: https://github.com/evilpegasus/APStats/blob/master/WA_School_Personnel_2018-2019.xlsx

2.1 Data Processing

With the raw salary data, all rows not containing a secondary teacher salary were filtered out and removed. Using a Python script, the secondary teacher salaries were read into a Pandas Dataframe, and the average of each district was found. To combine the datasets, the district names on the SAT dataset were renamed to match the naming convention of the salary spreadsheet. The renamed SAT dataset was then directly read into a Pandas Dataframe. The two Dataframes were merged by matching district names. Because not every district in the SAT dataset was listed on the salary dataset and not every salary district was listed on the SAT dataset, we only included districts with data from both sets.

3 Analysis

3.1 Unit 1

All statistical descriptions for both data sets were calculated with a computer. For the mean SAT data (see Figure 1), the distribution is roughly symmetric and unimodal. It is approximately normal-shaped. The median is at 1046.5 points and the mean is approximately 1043.995 points. The standard deviation is about 90.661 points. The minimum and maximum values are Taholah's low mean score of 803 points and Mercer Island's high mean score of 1316 points, which is an outlier (see Figure 2). The range is 513 points. The first quartile is 982 points and the third quartile is 1103.75 and the 1.5 * IQR is 182.625 points. This gives an upper fence of 1286.375 and a lower fence of 799.375. This confirms that Mercer Island's mean score of 1316 is an outlier and Taholah's mean of 803 points is not.
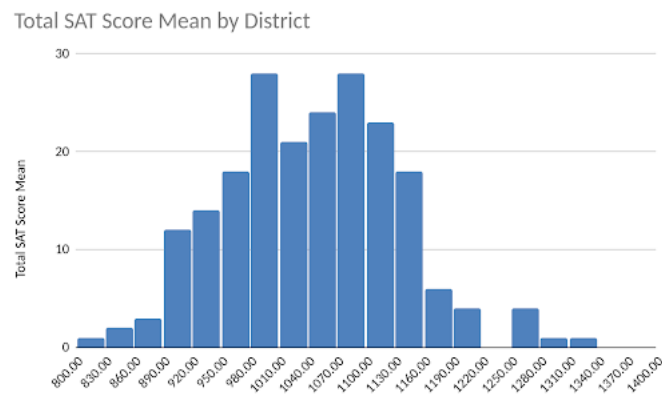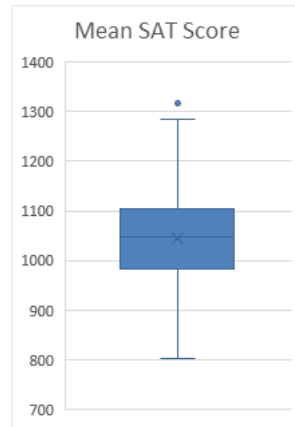

Figure 1: Histogram

Figure 2: Box and Whiskers Plot

For the average teacher salaries (see Figure 3), the data is multimodal and roughly symmetric and normal shaped. The mean is approximately $80,040.33 and the median is $80,158.9058. The standard deviation is about $9,241.6781. There are two outliers, $47,796.83333 (Coulee-Hartline) and $107,466.3848 (Everett). The range is about $59,669.5515. The first quartile is $73,999.66 and the third quartile is $85,988.53 and the 1.5 * IQR is $17,983.3. The fences are $56,016.36 and $103,971.8 which confirms that Coulee-Hartline and Everett are indeed outliers (see Figure 4).
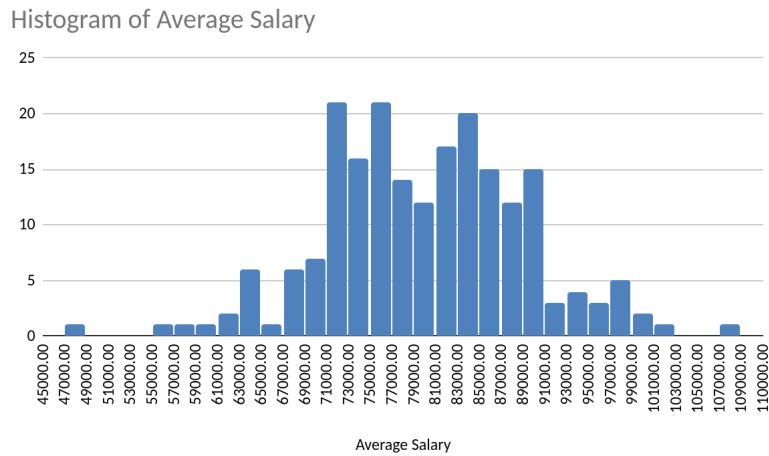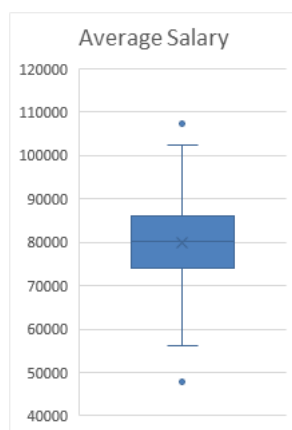

Figure 3: Histogram

Figure 4: Box and Whiskers Plot

3.2 Unit 2

Our scatter plot showed some positive correlation between teacher salary and mean SAT scores (see Figure 5). After plotting the residuals, we found the R^2 value to be R^2=0.09423098779302452. This R^2 value is small, but not negligible, and indicates that the linear relationship is weak. The R^2 value tells us that the model accounts for approximately 9.42% of the variance in average SAT scores by average teacher salaries. Some districts with interesting outliers in teacher salaries include Coulee-Hartline ($47,796.83) and Everett ($107,466.3848), the first district being a low outlier and the second being a high outlier. The district of Mercer Island (1316) is a high outlier in terms of average SAT score. Some lurking variables that might have adverse effects on our data and analysis include demographics, urban vs rural areas, and school district funding. For instance, in urban areas, there are more opportunities to enroll in SAT prep programs or get private tutoring than in rural areas. Likewise, the demographics within each school affects the population's abilities to further their SAT prep (through programs or tutoring) outside of school. School district funding positively correlates to SAT scores
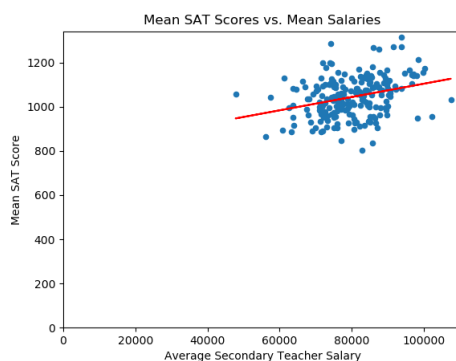

Figure 5: Scatterplot and Linear Regression Line

4

Linear Regression Equation

$$\hat{y} = 0.00301137x + 802.96409135$$

For each dollar increase in average secondary teacher salary, the mean SAT score of the district is predicted to increase by 0.00301137 points.

Our regression model predicts that a district paying teachers an average of $0 will have a mean SAT score of 802.96409135. This situation is not realistic as no teacher would work for free. Additionally, a prediction at the y-intercept significantly extrapolates beyond our dataset: the lowest mean salary reported was $47,796.83.

The R^2 value and the regression equation coefficients were both computed with the Scikit-learn software library for Python.

3.3 Unit 4

The Renton School District has 91 secondary teachers with a 6-figure salary (≥ 100000) out of a total of 295 secondary teachers. Given a simple random sample of 295 teachers from the set of all reported secondary teachers, what is the chance that more than 91 teachers are selected with six-figure salaries? There are 7234 secondary teachers in Washington State with six-figure salaries and 25511 total reported secondary teachers.

We will use a binomial distribution to model this. We first check the conditions:

$$np = 82.610 \geq 10$$
$$nq = 212.4 \geq 10$$
$$n = 295 < 0.10 * 25511$$

We can reasonably assume that most randomly selected salaries will be independent of each other. Using the binomial cumulative distribution function, we can find the probability that a random sample has equal to less than 91 teachers with six-figure salaries. The complement of that output will be our answer.

$$\sum_{i=0}^{x} \binom{n}{i}(p)^i(1-p)^{n-1} = 0.8447$$
$$1 - 0.8447 = 0.1553$$

Therefore, the probability that a randomly selected sample of 295 teachers has more than 91 teachers with a six-figure salary is 15.53%. This means that the Renton School District has a relatively high concentration of high-earning secondary teachers.

4 Conclusion

This study focused on if teacher salaries and student performance have a direct association and the possible factors involved. From this study, we can conclude that there was a positive association between teacher salaries and student performance in SAT scores

during the 2018-2019 school year. In context, if a district pays their secondary teachers more, their average SAT scores are likely to be higher.

The lurking variables that we theorized were demographics, urban vs rural areas, and school district funding, but it would be impractical to attempt to lessen the influence of these factors in our model. Similarly, another external factor is the variation in the cost of living between districts, which may have facilitated the differences in the demographics and teacher salaries of each school. For further research in the future, we could gather and use data from other states in the US similar to the data used in this study. We could also further consider the lurking factors stated above. Ultimately, our study exposed the flaws of the American education system and the unfair advantages wealth may give to students.

5 Appendix

WA Sat Scores: https://www.k12.wa.us/student-success/support-programs/dual-credit-programs/exam-based-dual-credit

WA Salary Data: https://www.k12.wa.us/safs-database-files

Repository of raw data, processed data, and analysis scripts: https://github.com/evilpegasus/apstats

The following pages contain the processed and combined data that was used in the analysis.