Family Unit Makeup and Its Impact on School Test Scores

Tony Blodgett

MIS500 – Foundations of Data Analytics

Colorado State University – Global Campus

Dwight Davis

June 9, 2019

**Background**

A few studies have shown a strong correlation between households that encourage and provide educational support along with a positive home academic environment (Brown et all, 2008). I would like to investigate and see if the types of family units within a given school district has an impact on the annual testing scores for the students in those districts regardless of the academic support level at home. I have always been a firm believer that the student is the biggest contributing factor in their own success and I would like to begin to either support or refute my beliefs.

A student’s success (or failure) and the factors that have had the most impact have been researched and evaluated since schools have been in session. These often-researched items typically focus on things at the school like teaching techniques, testing techniques and school makeup. More often than not, the makeup of a student’s home or the situation they find themselves in is not measured (Roscigno, Tomaskovic-Devey, Crowley, 2006)

**Tools and Data Sets**

In an effort to determine if a correlation exists between a student’s household makeup and their testing scores, multiple public data sets were used in conjunction with the R programming language and the RStudio IDE. For visualization of the results, a combination of the internal graphics capabilities from within R coupled with some options from the ggpubr visualization package will be used. The public data sets used for this study were obtained from multiple resources from the State of Arizona. The data sets include the Arizona portion of the 2010 census containing family unit makeup numbers by incorporated place (Office of Economic Opportunity, 2010), a list of schools by incorporated place provided by the Arizona Department of Education (ADE) (Great School.org, 2019) and finally the AzMerit testing data by school for school year 2018 also provided by ADE (AZ Department of Education, 2019).

**Planned Analysis & Expected Results**

To determine if there was a correlation between the size and type of a household unit and a given individual’s test scores, three datasets were relied upon for answers. The combination of these three data sets allowed for the connection of household makeup and to AzMerit testing score results for a given school. With the data containing one independent variable (family unit), with dependent groups, and one dependent variable (test scores) a Paired T-test was in order. A Paired T-test result of .05 or under would have rejected the null hypothesis. The expected result was a p value of less than .05 thus rejecting the null hypothesis in favor of the alternate hypothesis.

**Hypothesis**

The idea that a student is so influenced by their surroundings that they are incapable of achievement on a personal level is hard to understand. Is there no direct correlation between the individual and how they do on their own anymore? This is a question that this project was going to answer.

Null Hypothesis – H0 – The results of a student’s testing is impacted, positively or negatively, by the make up of the family unit they are in.

Alternate Hypothesis – H1 – The results of a student’s testing is not impacted, positively or negatively by the make up of the family unit they are in.

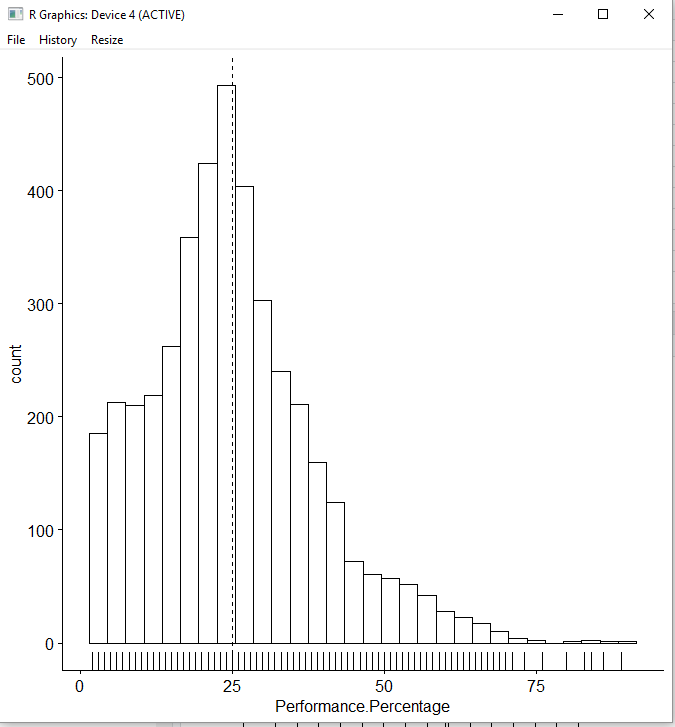
**Evaluation & Results**

As stated by Peterson (2003) the process of “weeding out and fixing or discarding inconsistent, incorrect or incomplete data is what is called data scrubbing” and is a necessary step in preparing data for use. In this instance, data scrubbing was needed to remove duplicates and inconstant data in both the 2010 census data as well as the list of schools by incorporated place data set.

In order to evaluate the data, some data reshaping was necessary. The AzMerit data in particular had to be reshaped to change the four column level data for each performance category and percentage into a two-column data set so that it could be evaluated. Once the data was transposed, it could be evaluated effectively.

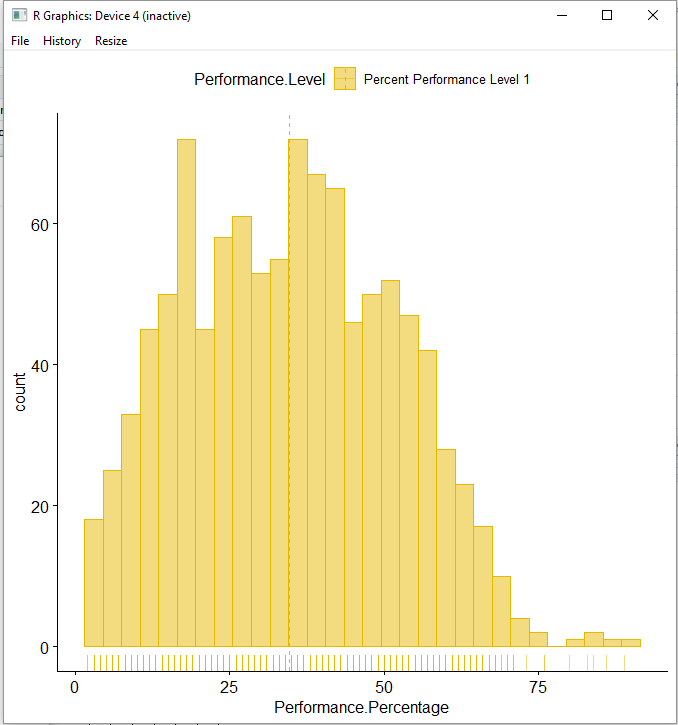
In an effort to have enough observations coupled with an adequate number of predictors (Seaman, Seaman & Allen, 2013) some initial filtering was done to eliminate abnormally small data values as well as to limit the observation to the Math only results.

When looking at the distribution of the results over the potential quartile results, the values were proven to be normalized where the mean of each results was about 25%. The graph below shows the results of the AzMerit percentage results.

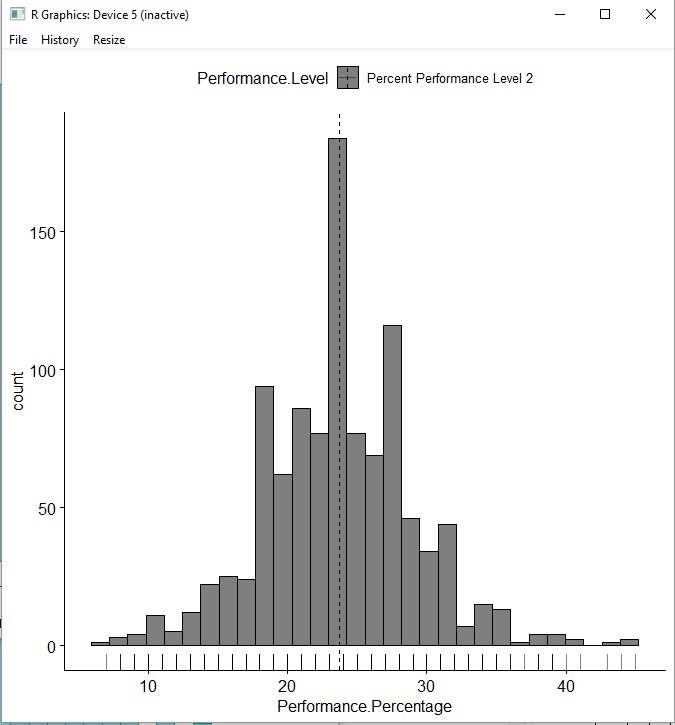


*Figure 1.* Overall Performance Percentage distribution by state

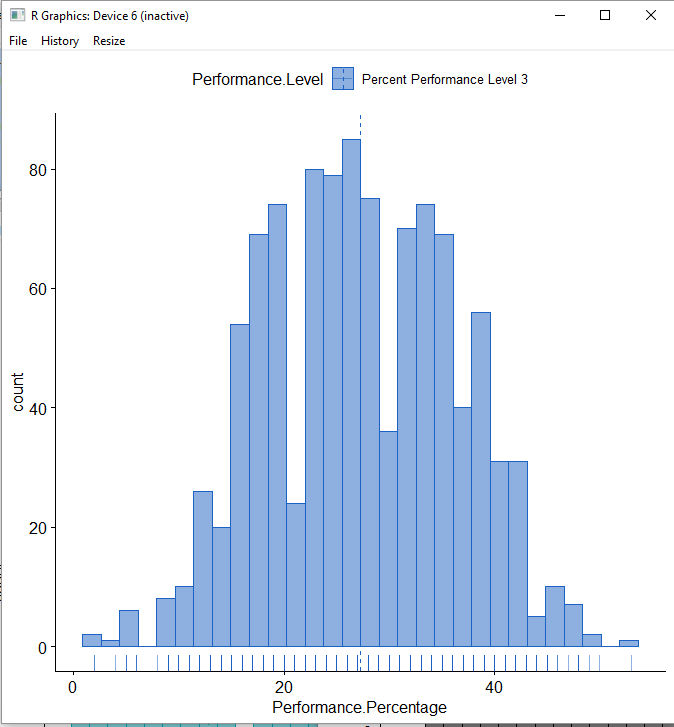
The next step was to check and see if the values for each percentage group was normalized as well. Each of the quartile groups did appear to be normalized as shown in the graphs below.



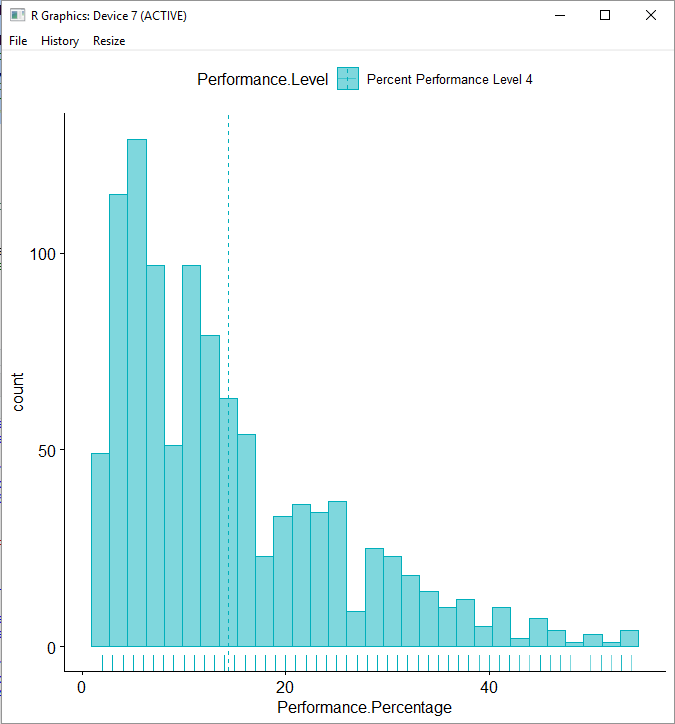
*Figure 2.* AzMerit Level 1 Performance Percentage test results for all schools



*Figure 3.* AzMerit Level 2 Performance Percentage test results for all schools



*Figure 4.* AzMerit Level 3 Performance Percentage test results for all schools



*Figure 5.* AzMerit Level 4 Performance Percentage test results for all schools

After checking summary statistics and looking for normalization, the process of evaluating the hypothesis was started. At this point it was discovered, after not being able to complete a valid statistical test, that a problem existed with either the data or its intended use. In the end, it was the intended use that was the problem. Upon further thought and investigation, the household unit numbers as well as the AzMerit test result percentages are both summary data sets with no real connection. Without some relationship between the data sets, there was no statistical method available for evaluation.

A conclusion on the hypothesis presented could have been reached if the actual test results were collected by participant versus summarized by school and if there was an indication, on a per test basis, what type of household unit type the test taker belonged to. Better evaluation and understanding of the underlying data, earlier in the process, should have revealed this flaw in reasoning.

In the event the required data was available to the group and was evaluated using the Paired T-test method I believe that the null hypothesis or H0 would have been rejected and the data would have shown that an individual’s household unit status does not have an impact on their testing results. Perhaps there is additional data compiled the Arizona Department of Education that can be requested in an effort to validate this opinion.

**References**

Roscigno, V., Tomaskovic-Devey, D., & Crowley, M. (2006). Education and the Inequalities of Place. Social Forces, 84(4), 2121-2145. Fabregas, K (2018).

Peterson, T. (2003). Data scrubbing. Computerworld, 37(6), 32. Retrieved from <https://csuglobal.idm.oclc.org/login?url=https://search-proquest-com.csuglobal.idm.oclc.org/docview/216095902?accountid=38569>

Seaman, Julia, Seaman, Christopher, & Allen, I. (2013). So Many Variables, So Few Observations. Quality Progress, 46(9), 34-41.

Sheridan, S. M., Warnes, E. D., Cowan, R. J., Schemm, A. V., & Clarke, B. L. (2004). Family-centered positive psychology: Focusing on strengths to build student success. Psychology in the Schools, 41(1), 7–17. <https://doi-org.csuglobal.idm.oclc.org/10.1002/pits.10134>

Office of Economic Opportunity (2010). Census 2010 & Census 2000 Demographics for Arizona Places [Data file]. Retrieved from <https://population.az.gov/census-data>

Bowen, G., Rose, R., Powers, J., & Glennie, E. (2008). The Joint Effects of Neighborhoods, Schools, Peers, and Families on Changes in the School Success of Middle School Students\*. Family Relations, 57(4), 504-516.

Great School.org (2019). Arizona School Districts [Data file]. Retrieved from <https://www.greatschools.org/schools/districts/Arizona/AZ/>

Arizona Department of Education (2019). Accountability and Research AzMerit, MSAA, AIMS and AZELLA Results [Data file]. Retrieved from <http://www.azed.gov/accountability-research/data/>