Potential Correlation of SAT Scores with Expenditure Per Pupil, SES, and IDEA

Kay Garbarino-Flowers

Seattle Pacific University

Part One

For part 1 we chose to look at the variables for current expenditure per pupil in average daily attendance in public elementary and secondary schools, percent of students eligible for free/reduced lunch (SES), and percent of students with disabilities (IDEA). All descriptive data were obtained using a Microsoft Office Excel worksheet called Descriptive.xls. This worksheet was provided by Prof. Mvududu. All data analyzed was provided within a worksheet called Project.xls. This worksheet was also provided by Prof. Mvududu.

Expenditure per Pupil in Average Daily Attendance in Public Elementary and Secondary Schools in the Academic Year of 2005 to 2006

Reviewing the histogram in Figure 1.1 for current expenditures per pupil in average daily attendance in public elementary and secondary schools in the academic year of 2005 to 2006 shows a positively skewed distribution with a skew value of 1.16 for the sample of all four regions. The following descriptives were obtained for the four regions considered as one sample. The mean of the distribution is $10,327.87 expenditure per student. The median of the distribution is $9,805.00 expenditure per student. The standard deviation for this distribution is $2,502.20. The inter-quartile range for the four region sample is $2787 per student, with $8639 to $11426 as the values for the 1st and 3rd quartiles, respectively.

Referring to Figure 1.2 for the four regions of the sample from which this data is derived, we found one outlier in the South region, Washington D.C. The South region has a distribution mean of $9720.88, a median of $8975, a standard deviation of $266.95, and a positive skew of 2.51. The inter-quartile range is $1822 with 1st and 3rd quartile values of $8155 and $10008. The expenditure per student for Washington DC, the outlier, is $18,339.

Reviewing the data for the West region, the distribution mean is $9244.92, the median is $8928, and the standard deviation is $2024.64. This distribution is positively skewed with a skew value of 0.52. The inter-quartile range for this region is $1983 with 1st and 3rd quartile value of $7919 and $9902, respectively.

Figure 1.1. Current Expenditure per Pupil

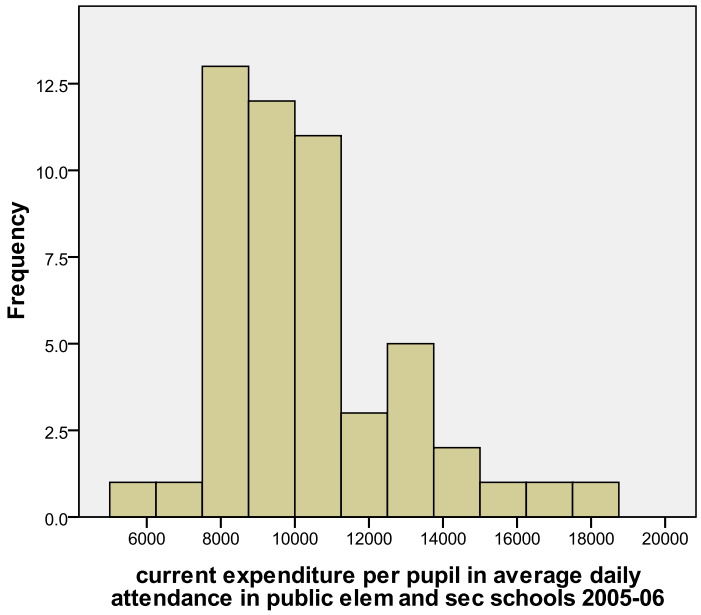

Figure 1.1. Displays current expenditure in dollar value per pupil in average attendance in public elementary and secondary schools for the 2005-2006 academic year. This histogram represents total expenditures that include West, Midwest, South, and Northeast regions.

Reviewing the data for the Midwest region, the distribution mean is $9905.42, the median is $10087, and the standard deviation is $814.05. This distribution is negatively skewed with a skew value of -0.91. The inter-quartile range for this region is $1179.75, with 1st and 3rd quartile values of $9374.25 and $10554, respectively.

The data for the Northeast region shows a positively skewed distribution with a skew value of 0.21. This distribution mean is $13601.44, the median is $13723, and the standard deviation is $1804.85. The inter-quartile range for the Northeast region is $2213, with 1st and 3rd quartile values of $12054 and $14277 respectively.

Figure 1. 2. Current Expenditure per Pupil for West, Midwest, South, and Northeast Regions

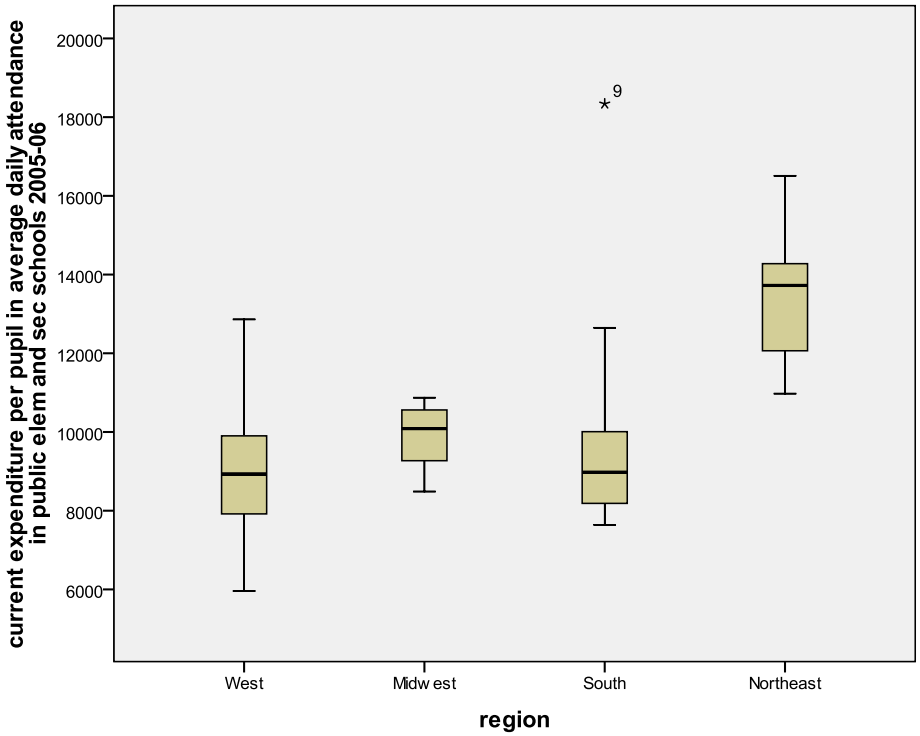

Figure 1.2. A box plot for West, Midwest, South, and Northeast regions that displays the current expenditures per pupil in average daily attendance in public elementary and secondary schools for the 2005-2006 academic year.

Reviewing the data for expenditure per student in thousands of dollars, we see the mean of $10,327.78 and a standard deviation of $2,502.20, creating a range of $7,825.58 to $12,829.98 between the mean and 1 standard deviation above and below the mean. Categorizing the data by region, three of the four regions, West, Midwest, and South, contain annual expenditures within

the above stated range. All three of these regions have mean and median values within this range. There is a second smaller frequency peak in the histogram between approximate values $12,500 and $13,750. The West region's highest annual expenditure is within this frequency range and the Midwest does not have any annual expenditure per student in this second higher frequency range. The South and Northeast regions have annual expenditures in this higher frequency range, although the South region's mean and median values fall below the values for the second frequency peak. The South region's outlier falls above this frequency range. The Northeast region's mean and median values fall into the second smaller frequency peak, which is between 1 and 2 standard deviations above the mean.

Percent Students Eligible for Free or Reduced Lunch in the Academic Year 2006-2007

Reviewing the histogram in Figure 1.3 for percent of SES shows a positively skewed distribution with a skew value of 0.61 for the sample of all four regions. The median of the distribution is 37.5% students, the mean of the distribution is 40.1% students, and the mode of this distribution is 40.7% students. The standard deviation for this distribution is 10.57%. The inter-quartile range for the four region sample is 15.7%, with 32.1% to 47.8% as the values for the 1st and 3rd quartiles, respectively.

Reviewing Figure 1.4 for the four regions of the sample from which this data is derived, we found three outliers. One outlier is shown in the West region, New Mexico, and two outliers are in the Northeast, New Hampshire and New York.

Let's look at the West region data first. Reviewing the data for the West region, the distribution mean is 40.49% SES, the median is 37.2%, and the mode is 40.7%. The standard deviation of this distribution is 8.37%. This distribution is positively skewed with a skew value

of 1.66. The inter-quartile range for this region is 5.3% with 1st and 3rd quartile value of 35.95% and 41.25%, respectively. The range overall is 29.7%; this is the second largest range for the statistic of percentage of SES. The box plot in Figure 1.4 reflects this data. New Mexico has a 60.9% rate of SES. Wyoming has 29.4% SES.

Figure 1.3. % of Students Eligible for Free or Reduced Lunch (SES)

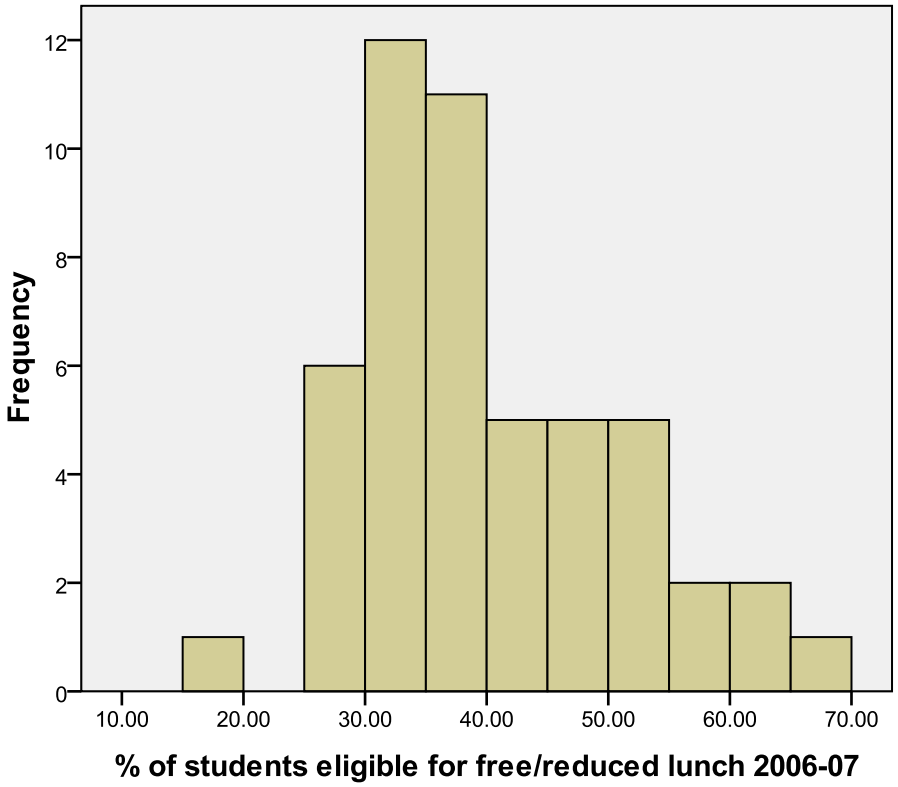

Figure 1.3. Histogram showing percent of SES in the 2006-2007 academic year.

Reviewing the SES data for the Midwest region, the distribution mean is 34.42%, the median is 34.9%, and the standard deviation is 3.77%. The mode is not applicable for this region. This distribution is negatively skewed with a skew value of -0.05. The inter-quartile range for this region is 6.65%, with 1st and 3rd quartile values of 30.9% and 37.55%, respectively.

There are no outliers in this region for this variable. The Midwest has a tight range for percent SES with South Dakota at 28.9% and Kansas at 39.8%.

Figure 1.4. % of Students Eligible for Free or Reduced Lunch (SES)

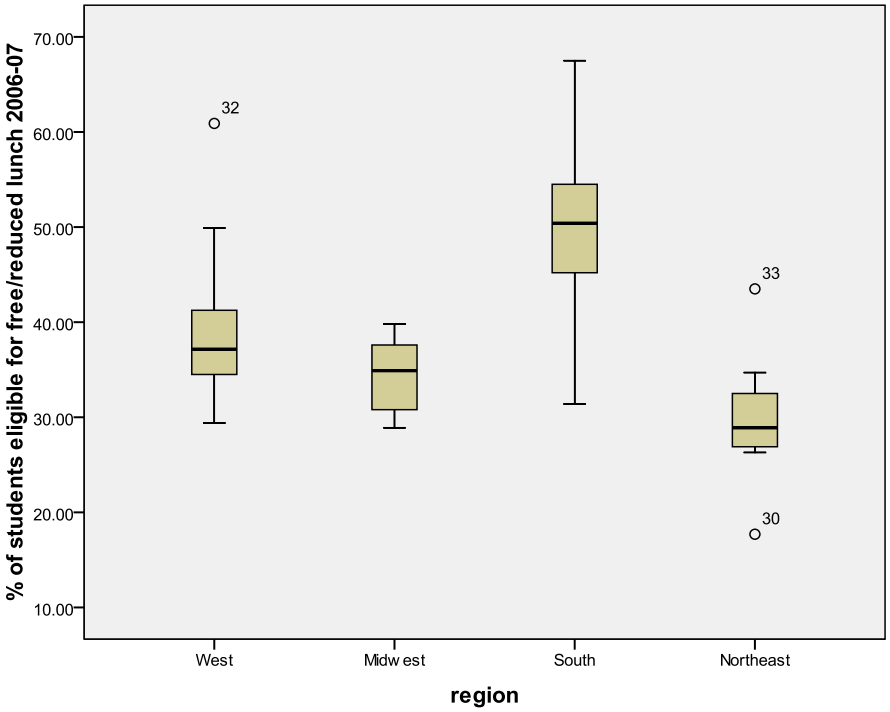

Figure 1.4. Box plot for West, Midwest, South, and Northeast regions showing the percent of SES in the 2006-2007 academic year.

Reviewing the SES data for the South region, the distribution mean is 49.14%%, the median is 50.4%, and the standard deviation is 9.53%. The mode is not applicable for this region. This distribution is negatively skewed with a skew value of -0.24. The inter-quartile range for this region is 9.3%, with 1st and 3rd quartile values of 45.2% and 54.5%, respectively. There are no outliers in this region for this variable. The South region has a large range for percent SES with Virginia at 31.4% and Mississippi at 67.5%.

Reviewing the SES data for the Northeast region, the distribution mean is 29.78%, the median is 28.9%, and the standard deviation is 7.02%. The mode is not applicable for this region. This distribution is positively skewed with a skew value of 0.41. The inter-quartile range for this region is 5.6%, with 1st and 3rd quartile values of 26.9% and 32.5%, respectively. There are two outliers in this region for the SES variable. Excluding the outliers, the Northeast region has a tight range for percent SES with Vermont at 26.3% and Maine at 34.7%. The outliers for this region are New Hampshire with 17.7% SES. New York has 43.5% SES. These outliers explain the large standard deviation while the mean and median values are within 0.88% of each other.

Reviewing the data for the frequency of percentage of students that are eligible for free/reduced lunch, we see the mean at 40.1% with a standard deviation of 10.57%. This data creates a range of 29.53% to 50.67% that is 1 standard deviation above and below the mean. The histogram shows a frequency peak that falls between the mean and 1 standard deviation below the mean. Incidence of students that are eligible for free/reduced lunch continues at a rate that pulls this distribution toward 70%, more than 2 standard deviations above the mean. This produces this distribution's positive skew. Looking at the peak of the distribution, all four regions contain states with a student population that is eligible for free/reduced lunch between the mean and 1 standard deviation above and below the mean. The West and Midwest regions have mean and median values between the mean and 1 standard deviation below the mean. The frequency of students eligible for free/reduced lunch at the peak of the histogram distribution to the positive tail of this distribution fall between the mean and 2 standard deviations above the mean. 100% of the South region falls into this aforementioned area below the distribution curve. Excluding its outlier, New York, the Northeast region's data falls between the mean and 2

standard deviations below the mean. Overall, the West and Midwest regions fall into the middle of the distribution for percentage of students eligible for free/reduced lunch, the South region falls into the higher end of the distribution for percentage of students eligible for free/reduced lunch. The Northeast region, excluding its outlier of New York, falls into the lower end of the distribution for percentage of students eligible for free/reduced lunch.

Percent Students with Disabilities in the Academic Year 2006-2007

Figure 1.5. % of Students with Disabilities (IDEA)

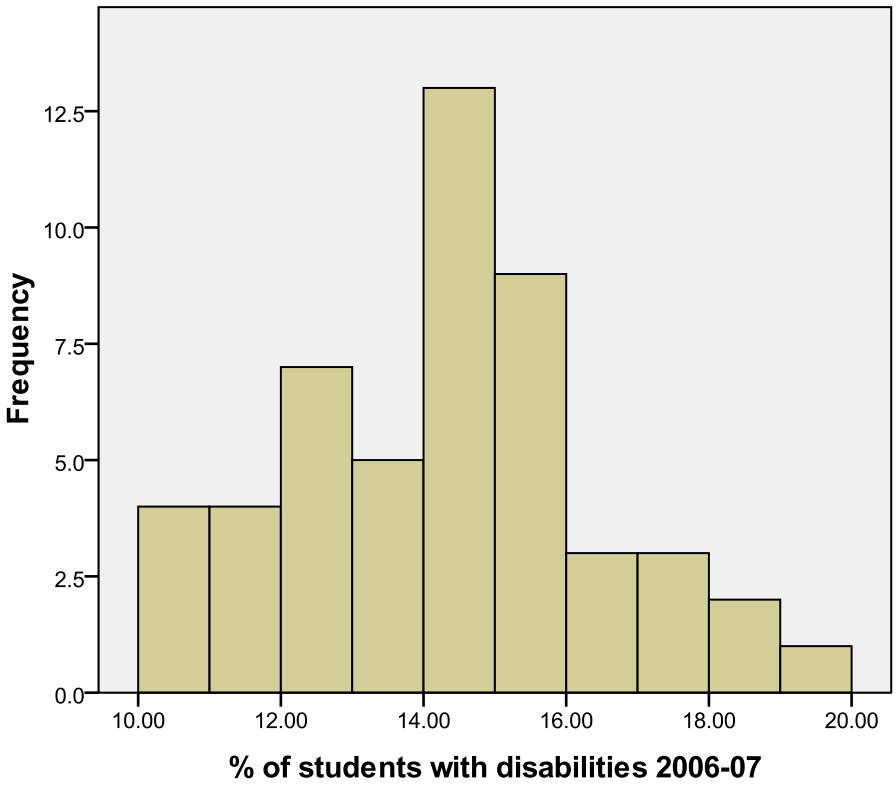

Figure 1.5. Histogram shows the IDEA percentage for regions West, Midwest,, South, and Northeast in the 2006-2007 academic year.

Reviewing the histogram in Figure 1.5 for percent of students with disabilities (IDEA) in 2006-07 shows a positively skewed distribution with a skew of 0.23 for the sample of all four regions. The mean of the distribution is 14.22% students. The median of the distribution is 14.3% students. The mode of this distribution is 14.7% students. The standard deviation for this distribution is 2.13%. The inter-quartile range for the four region sample is 3.0%, with 12.4% to 15.4% as the values for the 1st and 3rd quartiles, respectively. Reviewing the data displays no outliers for the variable of percent students with disabilities for 2006-2007 within any of the four regions.

Let's look at the West region data first. Reviewing the IDEA data for the West region, the distribution mean is 12.41% students with disabilities, the median is 11.9%, and the mode is 10.5%. The standard deviation of this distribution is 1.75%. This distribution is positively skewed with a skew value of 1.03. The inter-quartile range for this region is 2.0% with 1st and 3rd quartile value of 11.4% and 13.4%, respectively. The box plot in Figure 1.6 reflects this data. New Mexico has a 14.6% rate of IDEA. Wyoming has 16.4% IDEA. These two states within the West region produce the positive skew of this region's distribution.

Reviewing the IDEA data for the Midwest region, the distribution mean is 14.9%, the median is 14.7%, and the mode is 14.7%. The standard deviation for the Midwest region's distribution is 0.88%. This distribution is positively skewed with a skew value of 1.43. The inter-quartile range for this region is 1.15%, with 1st and 3rd quartile values of 14.25% and 15.40%, respectively. The box plot in Figure 1.6 reflects this data. The Midwest has a tight range for percent students with disabilities; Kansas has 14.00% students with disabilities, Indiana has 17.10%. This produces an overall range of 3.1%.

Figure 1.6. % of Students with Disabilities

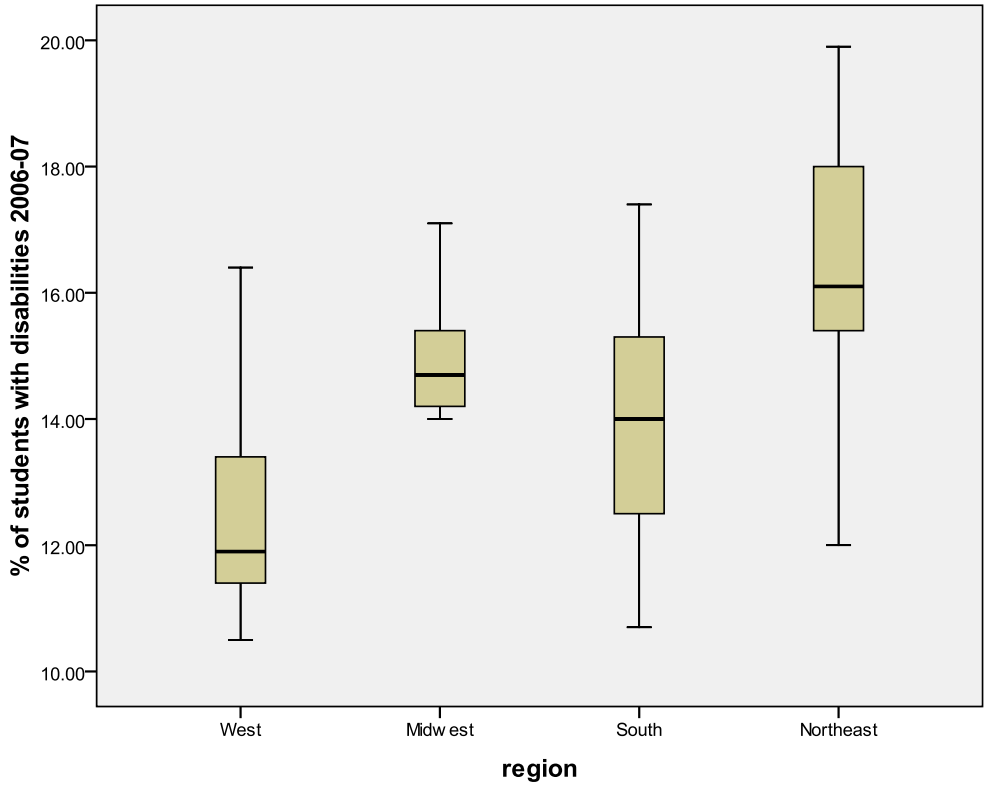

Figure 1.6. Box plot of percent of students with disabilities within West, Midwest, South, and Northeast regions in the 2006-2007 academic year.

Reviewing the IDEA data for the South region, the distribution mean is 13.99%, the median is 14.0%, and the mode is 15.3%. The South region standard deviation for percent students with disabilities is 1.75%. This distribution is almost perfect with a skew value of 0.01. The inter-quartile range for this region is 2.8%, with 1st and 3rd quartile values of 12.5% and 15.3%, respectively. The box plot in Figure 1.6 reflects this data. The South region has the second largest range for percent students with disabilities with Texas at 10.7% and West Virginia at 17.4%.

Reviewing the IDEA data for the Northeast region, the distribution mean is 16.34%, the median is 16.1%, and the standard deviation is 2.31%. The mode is not applicable for this region. This distribution has a slight negative skew with a skew value of -0.39. The inter-quartile range for this region is 2.6%, with 1st and 3rd quartile values of 15.4% and 18%, respectively. The box plot in Figure 1.6 reflects this data. The Northeast region has the largest range for percent students with disabilities with Connecticut at 12.0% and Rhode Island at 19.9%.

Looking at the frequency of percentage of students with disabilities, the mean of the four region distribution is 14.21% with a standard deviation of 2.13%, creating a range of 12.08% to 16.34% that is within 1 standard deviation from the mean. Review of the above box plot in Figure 1.6 shows that the West region falls within the normal range to more than 1 standard deviation below the mean. The Midwest region falls predominantly between the mean and 1 standard deviation above the mean. The South region covers the distribution from more than 1 standard deviation above and below the mean with its inter quartile range falling between the mean and 1 standard deviation above and below the mean. The Northeast region's data falls predominantly between 1 and 2 standard deviations above the mean for the overall distribution.

Part Two

For part two, values were obtained from Part 2 Data Ouput provided on Blackboard by Prof. Mvududu. Three variables were analyzed in Part 2: current expenditure per pupil in average daily attendance in public elementary and secondary schools for 2005-2006 academic year, percent students eligible for free/reduced lunch for 2006-2007 academic year, and percent of students with disabilities 2006-2007 academic year.

Current Expenditure Per Pupil in Average Daily Attendance in Public Elementary and Secondary Schools in the Academic Year 2005-06

In order to see if there is a statistical and practical significance between the four regions for expenditure, Analysis of Variance, ANOVA was used.

When looking at Levene's test of equality of error for variances across groups, we see an F value of 1.45 and a P value of 0.24. Assuming the null hypothesis is true, this P value tells us there is a 24% chance of obtaining these results. In this case, our null hypothesis assumes that the error of variance is equal across the different groups, or assumes homogeneity across groups. With an F value of 1.45, we can assume the error of variance is equal across groups and is not due to chance.

Tests Between Subjects Effects shows F=9.75 which is greater than $F_{(3,47)}=4.24$, $p<0.01$. The Tests Between Subjects Effects F ratio tells us there is a significant difference between regions for the expenditures variable since it is larger than the $F_{(3,47)}$ value. The effects size, eta squared, is 0.38, a strong effect.

In order to see where the variance lies between regions, Tukey's HSD was used. Setting the alpha level at 0.05, there is significance in the Mean Difference between the Northeast region and the Midwest, South, and West regions. Multiple comparisons failed to show significance in the Mean Differences between the West, Midwest, and South regions when compared to one another. See Table 2.1 for results of Multiple Comparisons placed in matrix form.

Table 2.1. Multiple Comparisons Results for Current Expenditure Per Pupil

Region	West	Midwest	Northeast	South
West	--------	-660.49	-4356.52*	-475.96
Midwest	660.49	--------	-3696.03*	184.53
Northeast	4356.52*	3696.03*	--------	3880.56*
South	475.96	-184.53	-3880.56*	--------

Table 2.1. Results of multiple comparisons placed in matrix form, $P < 0.05$ for current expenditure per pupil in average daily attendance in public elementary and secondary schools 2005-06. * indicates significance.

As seen in part 1, the Northeast region has greater expenditure per pupil than does the Midwest, South, and West regions, excluding the South region's outlier. The Northeast region's expenditure per pupil falls between 1 and 2 standard deviations above the mean, whereas the remaining regions' expenditure per pupil fall between the mean and 1 standard deviation below the mean. This may explain the significance in Mean Differences between the Northeast region and the remaining regions. Eta squared value of 0.38 showed a strong effect. The strong effect indicates there is a practical significance in the mean differences as well as a statistical significance.

Percent of Students Eligible for Free/Reduced Lunch in the Academic Year 2006-07

In order to see if there is a statistical and practical significance between the four regions for SES, ANOVA was used.

Four regions were compared for percent of students eligible for free/reduced lunch in the 2006-07 academic year. These regions were West with an N of 12, Midwest with an N of 12, South with an N of 17, and Northeast with an N of 9.

When looking at Levene's test of equality of error for variances across groups, we see an F value of 1.30 and a P value of 0.29. Assuming the null hypothesis is true, this P value tells us there is a 29% chance of obtaining these results. In this case, our null hypothesis assumes that the error of variance is equal across the different groups, or assumes homogeneity across groups. With an F value of 1.30, we can assume the error of variance is equal across groups and is not due to chance.

Tests Between Subjects Effects shows F = 14.87 with $F_{(3,46)} = 4.24$, $p<0.01$. This F ratio tells us there is a significant difference between regions for the expenditures variable since F + 14.87 is larger than $F_{(3,46)}$. The effects size, eta squared, is 0.49, a strong effect.

In order to see where the variance lies between regions, Tukey's HSD was used. Setting the alpha level at 0.05, there is statistical significance in the Mean Difference between the West region and the South and Northeast regions. Continuing the comparisons between individual regions, there is statistical significance in the Mean Difference between the Midwest and South regions. Comparing the South to all other regions shows statistical significance in the Mean Difference. Comparing the Northeast, there is a significant Mean Difference between this region and the West and South regions. See the matrix below in Table 2.2 for a matrix of the Mean Difference between the regions.

Tukey's HSD failed to show a statistical difference between the West and Midwest regions. Review of the box plot in Figure 2.1 shows that the IQR of each region falls between the mean and 1 standard deviation below the mean, supporting the failure to show a statistical difference. The matrix of Tukey's HSD results and the box plot display significant difference between the mean of the South region and the remaining regions.

Table 2.2. Multiple Comparisons Results for % Students Eligible for Free/Reduced Lunch

Region	West	Midwest	Northeast	South
West	--------	5.14	9.79*	-9.57*
Midwest	5.14	--------	4.65	-14.71*
Northeast	-9.79*	-4.65	--------	-19.38*
South	9.57*	14.71*	19.36*	--------

Table 2.2. Results of multiple comparisons placed in matrix form, P<0.05 for percents students eligible for free or reduced lunch.

Figure 2.1. % Students Eligible for Free/Reduced Lunch

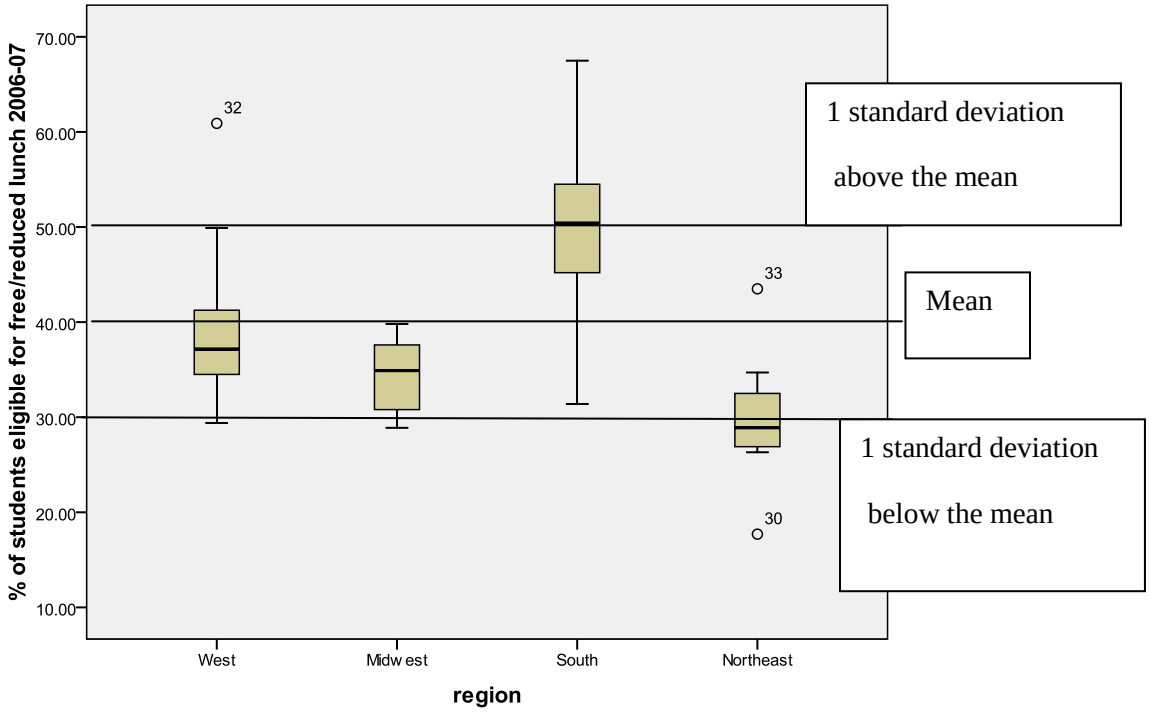

Figure 2.1. Box plot of percent of students eligible for free or reduced lunch with markers for all regions' mean and standard deviation.

Percent of Students With Disabilities in the Academic Year 2006-07

In order to see if there is a statistical and practical significance between the four regions for IDEA, ANOVA was used.

Four regions were compared for % of students with disabilities in the 2006-07 academic year. These regions were West with an N of 13, Midwest with an N of 12, South with an N of 17, and Northeast with an N of 9.

When looking at Levene's test of equality of error for variances across groups, we see an F value of 2.55 and a P value of 0.07. Assuming the null hypothesis is true, this P value tells us there is a 7% chance of obtaining these results. In this case, our null hypothesis assumes that the error of variance is equal across the different groups, or assumes homogeneity across groups. With an F value of 2.55, we can assume the error of variance is equal across groups and is not due to chance.

Tests Between Subjects Effects shows an $F = 10.25$ with $F_{(3,47)} = 4.24$, $p < 0.01$. This F ratio tells us there is a significant difference between regions for the expenditures variable since $F = 10.25$ is greater than the value for $F_{(3,47)}$. The effects size, eta squared, is 0.40, a strong effect.

In order to see where the variance lies between regions, Tukey's HSD was used. Setting the alpha level at 0.05, there is statistical significance in the Mean Difference between the West region and the Midwest and Northeast regions. This is supported by the box plot and histogram seen in Figures 1.5 and 1.6 in part 1, which shows the West's IQR below the mean and spanning into 1 standard deviation below the mean. Continuing the comparisons between individual regions, there is statistical significance in the Mean Difference between the Midwest and West

regions. This is supported by the box plot and histogram seen in Figures 1.5 and 1.6 in part 1, which show the Midwest's IQR between the mean and 1 standard deviation above the mean compared to the West's IQR which falls below the mean. Comparing the South to the Northeast region shows statistical significance in the Mean Difference. The box plot in Figure 1.6 displays the Northeast's IQR spanning 1 standard deviation above the mean, while the South region displays a normal distribution with the South's mean at the mean of all four regions. Comparing the Northeast, there is a significant Mean Difference between this region and the West and South regions. I believe this is due to the IQR's of West and South regions either spanning the total mean or beginning below the total mean and spanning 1 standard deviation below the mean. See the matrix in Table 2.3 below for the Mean Difference between the regions.

Table 2.3. Multiple Comparisons Results for % Students with Disabilities

Region	West	Midwest	Northeast	South
West	--------	-2.49*	-3.94*	-1.58
Midwest	2.49*	--------	-1.4	0.91
Northeast	3.94*	1.44	--------	2.36*
South	1.58	-0.91	-2.36*	--------

Table 2.3. Results of multiple comparisons placed in matrix form, $P<0.05$ for percent students with disabilities.

Part Three

I selected expenditure and SAT scores, SES and SAT scores, and IDEA and SAT scores as the three pairs for analysis.

Expenditure per pupil and SAT scores

Analyzing expenditure and SAT scores reveals a negative correlation between the expenditures per student and SAT scores. The Correlation Coefficient ranges from -0.39 to -0.42 for math, reading, and writing SAT scores and the scores' correlation with expenditures per pupil. The critical value for r is 0.284 with 49 degrees of freedom and an alpha level set at 0.05. According to Guilford's suggested interpretation, the correlation coefficients suggest a moderate and substantial relation. If we look at r^2, the coefficient of determination ranges from 0.15 to 0.17. This indicates to what degree SAT scores are correlated with expenditure per pupil. See Table 3.1 for a summary of correlation coefficients and coefficient of determination for expenditure and SAT scores.

Table 3.1. Correlation Coefficients and Coefficient of Determination for Expenditure and SAT Scores

	Math SAT	Reading SAT	Writing SAT
r	-0.39	-0.42	-0.40
r^2	0.15	0.17	0.16

Table 3.1. Summarizes the Correlation Coefficient (r) and Coefficient of Determination (r^2) when looking at the relationship between expenditures per pupil with average daily attendance within elementary or secondary school for 2005-2006 academic year and SAT scores for math, reading, and writing.

The corresponding regression equations are as follows for analysis of this pair of variables.

Expenditures and Reading SAT scores: $y = -0.0063x + 599.76$

Expenditures and Math SAT scores: $y = -0.0059x + 601.42$

Expenditures and Writing SAT scores: $y = -0.006x + 587.02$

For the expenditures independent variable and three separate SAT dependent variables, these equations indicate a negative slope of -0.0059 to -0.0063 and a y-intercept between 587.02 and 601.42 when x equals 0.

The above data tell me that there may be a moderate statistical correlation between expenditure per pupil and SAT scores. However, the Coefficient of Determination reveals that the independent variable, expenditure, has 15% to 17% information about the dependent variable, SAT scores. Therefore, I believe the practicality of using this correlation is low.

Students eligible for free or reduced lunch and SAT scores

The analysis for the variables of students eligible for free or reduced lunch (SES) and SAT scores has a critical value for r of 0.279 with 48 degrees of freedom and an alpha level set at 0.05. Referring to Table 3.2, there does not appear to be a correlation between this pair of variables because the Correlation Coefficient (r) does not meet the critical value for any of the SAT score variations.

Table 3.2. Correlation Coefficients and Coefficient of Determination for SES and SAT Scores

	Math SAT	Reading SAT	Writing SAT
r	-0.08	0.02	0.08
r^2	0.01	0.00	0.01

Table 3.2. Summarizes the Correlation Coefficient (r) and Coefficient of Determination (r^2) when looking at the relationship between students eligible for free or reduced lunch for 2006-2007 academic year and SAT scores for math, reading, and writing.

The corresponding regression equations are as follows for analysis of this pair of variables.

SES and Reading SAT scores: $y = 0.0852x + 532.29$

SES and Math SAT scores: $y = -0.2916x + 552.85$

SES and Writing SAT scores: $y = 0.286x + 514.87$

For the SES independent variable and three separate SAT dependent variables, these equations indicate a slope of -0.2916 to 0.286 and a y-intercept between 514.87 and 552.85 when x equals 0.

Students with disabilities and SAT scores

When looking at the correlation between students with disabilities (IDEA) and SAT scores, we find the critical value of r to be 0.279 for 49 degrees of freedom and an alpha level set

at 0.05. Reviewing Table 3.3 below, we see that the Correlation Coefficient for any of the three pairs did not meet the critical value of r.

Table 3.3. Correlation Coefficients and Coefficient of Determination for IDEA and SAT Scores

	Math SAT	Reading SAT	Writing SAT
r	-0.07	-0.06	-0.05
r^2	0.00	0.00	0.00

Table 3.3. Summarizes the Correlation Coefficient (r) and Coefficient of Determination (r^2) when looking at the relationship between students with disabilities for 2006-2007 academic year and SAT scores for math, reading, and writing.

The corresponding regression equations are as follows for analysis of this pair of variables.

IDEA and Reading SAT scores: $y = -1.1568x + 551.39$

IDEA and Math SAT scores: $y = -1.271x + 558.66$

IDEA and Writing SAT scores: $y = -0.9371x + 538.69$

For the IDEA independent variable and three separate SAT dependent variables, these equations indicate a negative slope of -0.9371 to -1.271 and a y-intercept between 538.69 and 558.66 when x equals 0.

The above analysis was performed on data that spanned one academic year per set of variables. In order to verify the small correlation between expenditure per pupil and SAT scores,

it would be necessary to perform this analysis over a several year period to determine if there are trends in correlation.