***Model 1.)***

The REG Procedure
Model: model1
Dependent Variable: testscr

| | |
|---|---|
| Number of Observations Read | 420 |
| Number of Observations Used | 420 |

Analysis of Variance

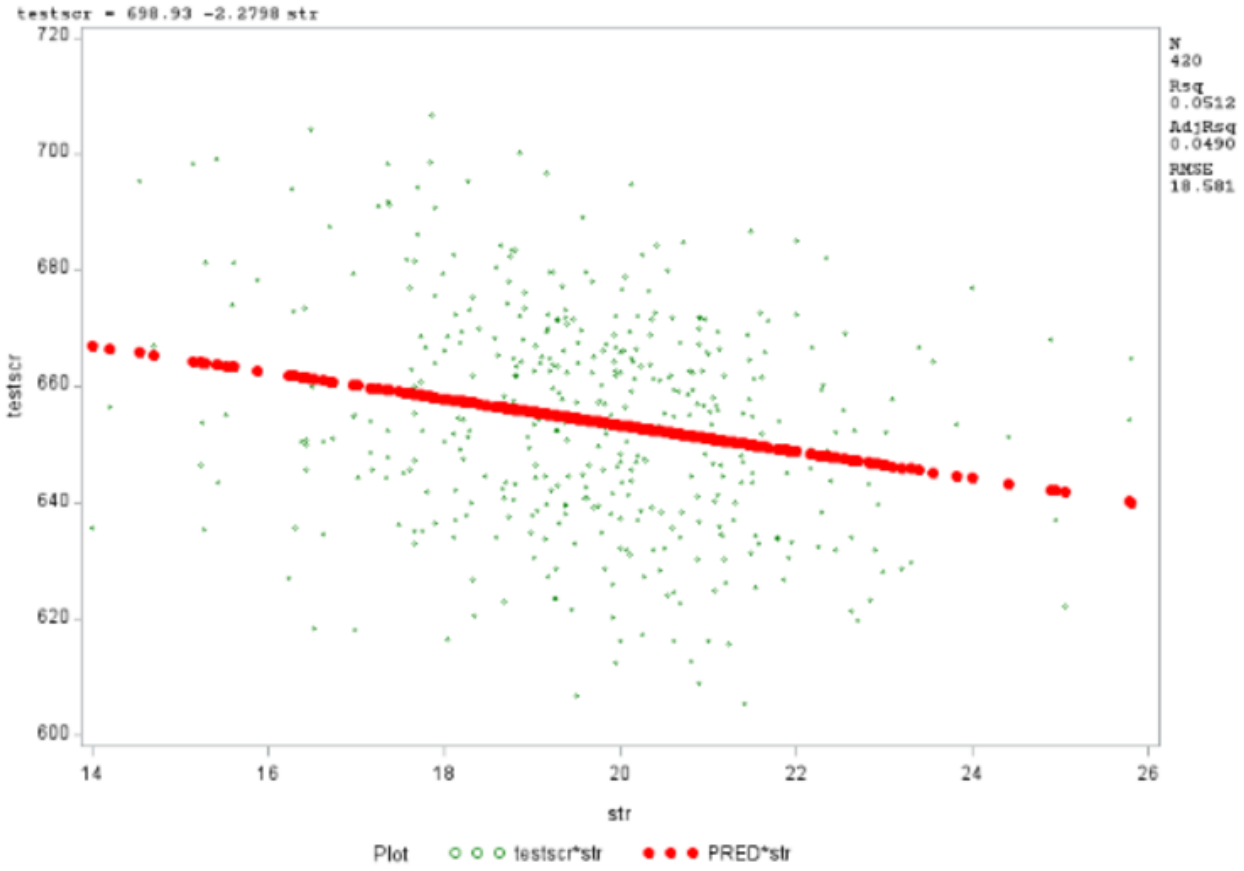
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 1 | 7794.11005 | 7794.11005 | 22.58 | <.0001 |
| Error | 418 | 144315 | 345.25235 | | |
| Corrected Total | 419 | 152110 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 18.58097 | R-Square | 0.0512 |
| Dependent Mean | 654.15655 | Adj R-Sq | 0.0490 |
| Coeff Var | 2.84045 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 698.93295 | 9.46749 | 73.82 | <.0001 |
| str | 1 | -2.27981 | 0.47983 | -4.75 | <.0001 |

- Based on the p-value of <.0001, you can see that the probability of obtaining a more extreme value is very small, so this variable is highly statistically significant at the 1% level.
- The model does not appear to fit the data well. This can be seen through the R^2 value of 0.0512. This is interpreted as 5.12% of the variation in Y is explained by variation in X, which is a very small amount. This could be due to the fact that many relevant predictor variables were left out of the regression. The adjusted R^2 is also very small and similar to R^2, which is expected since there's only one predictor variable.


testscr = 698.93 -2.2798 str
N
420
Rsq
0.0512
AdjRsq
0.0490
RMSE
18.581
720
700
680
660
640
620
600
testscr
14
16
18
20
22
24
26
str
Plot
testscr*str
PRED*str

***Model 2.)***

**The SAS System**

The REG Procedure
Model: model2
Dependent Variable: testscr

| | |
|---|---|
| Number of Observations Read | 420 |
| Number of Observations Used | 420 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 2 | 77801 | 38901 | 218.30 | <.0001 |
| Error | 417 | 74308 | 178.19690 | | |
| Corrected Total | 419 | 152110 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 13.34904 | R-Square | 0.5115 |
| Dependent Mean | 654.15655 | Adj R-Sq | 0.5091 |
| Coeff Var | 2.04065 | | |

**Parameter Estimates**

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 638.72915 | 7.44908 | 85.75 | <.0001 |
| str | 1 | -0.64874 | 0.35440 | -1.83 | 0.0679 |
| avginc | 1 | 1.83911 | 0.09279 | 19.82 | <.0001 |

- The results of this regression are a little surprising. For one, when looking at the p-value, class size (str) becomes statistically significant at the 10% level, but not the 5% level. This decrease could be due to multicollinearity. Also, we find that average income per capita (avginc) does have a statistically significant effect on Y at the 1% level. This can be seen through its low p-value of <.0001, which proves that the probability of getting a more extreme value than the null is very small, interpreted by the p-value of <.0001.
- This model is a much better fit of the data than the first model. This can be seen through the R^2 value of .5115, which means that 51.15% of the variation in Y is caused by variation in X. The Adjusted R^2 makes sense that it decreased the R^2 slightly as it accounted for the degrees of freedom in the model. The high F-Value also hints that the model is statistically significant.

***Model 3.)***

**The SAS System**

The REG Procedure
Model: model3
Dependent Variable: testscr

| | |
|---|---|
| Number of Observations Read | 420 |
| Number of Observations Used | 420 |

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 3 | 119757 | 39919 | 513.28 | <.0001 |
| Error | 416 | 32353 | 77.77151 | | |
| Corrected Total | 419 | 152110 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 8.81882 | R-Square | 0.7873 |
| Dependent Mean | 654.15655 | Adj R-Sq | 0.7858 |
| Coeff Var | 1.34812 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 685.03118 | 5.30955 | 129.02 | <.0001 |
| str | 1 | -0.83073 | 0.23426 | -3.55 | 0.0004 |
| avginc | 1 | 0.52727 | 0.08335 | 6.33 | <.0001 |
| meal_pct | 1 | -0.50631 | 0.02180 | -23.23 | <.0001 |

- Based on the p-value of <.0001, the variable that measures the percentage of students that qualify for a reduced price lunch is statistically significant at the 1% level, along with both of the other variables. The large t-value of -23.23 also provides evidence against the null hypothesis that there is a significant effect.
- The estimated coefficient for meal_pct in this model is -0.50631. This can be interpreted as for every 1% rise in the percentage of students that qualify for a reduced price lunch, the average test score decreases by 0.51 points.

***Model 4.)***

**The SAS System**

The REG Procedure
Model: model4
Dependent Variable: testscr

| | |
|---|---|
| Number of Observations Read | 420 |
| Number of Observations Used | 420 |

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 4 | 120368 | 30092 | 393.43 | <.0001 |
| Error | 415 | 31742 | 76.48558 | | |
| Corrected Total | 419 | 152110 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 8.74560 | R-Square | 0.7913 |
| Dependent Mean | 654.15655 | Adj R-Sq | 0.7893 |
| Coeff Var | 1.33693 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 678.35543 | 5.77062 | 117.55 | <.0001 |
| str | 1 | -0.63969 | 0.24194 | -2.64 | 0.0085 |
| avginc | 1 | 0.52188 | 0.08268 | 6.31 | <.0001 |
| meal_pct | 1 | -0.49940 | 0.02176 | -22.96 | <.0001 |
| comp_stu | 1 | 19.84323 | 7.01823 | 2.83 | 0.0049 |

- The model does the best job at fitting our data. Its R^2 value of 0.7913 shows that 79.13% of the variation in Y is caused by the variation in X. This is a huge improvement from our first and second model, but not much more than the third model that has a R^2 of .7873.
- The p-value on the computers per student (comp_stu) shows that it is statistically significant at a 1% level.

***Complete Regression Output Summary***

Key: *** - statistically significant at the 1% level
* - statistically significant at the 10% level

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| **intercept** | 698.93 | 638.73 | 685.0312 | 678.3554*** |
| **s.e.** | 9.467 | 7.4491 | 5.3096 | 5.7706 |
| **str** | -2.2798*** | -0.6487* | -0.8307*** | -0.6397*** |
| **s.e.** | 0.4798 | 0.3544 | 0.2343 | 0.2419 |
| **avginc** | | 1.8391*** | 0.5273*** | 0.5219*** |
| **s.e.** | | 0.0928 | 0.0834 | 0.0827 |
| **meal_pct** | | | -0.50631*** | -04994*** |
| **s.e.** | | | 0.0218 | 0.0218 |
| **comp_stu** | | | | 19.8432*** |
| **s.e.** | | | | 7.0182 |
| **R2** | 0.0512 | 0.5115 | 0.7873 | 0.7913 |
| **Adj R2** | 0.0490 | 0.5091 | 0.7858 | 0.7893 |

***Method of Semi-Averages:***

The SAS System

The MEANS Procedure

| split | N Obs | Variable | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|---|---|
| 1 | 210 | str | 210 | 18.1867016 | 1.2395347 | 14.0000000 | 19.7010860 |
| | | testscr | 210 | 658.2783319 | 19.6691683 | 606.7500000 | 706.7500000 |
| 2 | 210 | str | 210 | 21.0941489 | 1.1797176 | 19.7453308 | 25.7999992 |
| | | testscr | 210 | 650.0347642 | 17.5151403 | 605.5500488 | 694.8000488 |

***Calculate SSR by hand:***

Below are the calcluations used in my excel document to find the SSR

| Obs | ob | district | split | str | testscr | | | | Predicted | Residuals | Squared Residuals | SSR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | McCloud Union Elementary | 1 | 14 | 635.5999756 | | X(bar) first split | | 670.149006 | -34.54903037 | 1193.6355 | 144778.2581 |
| 2 | 2 | Chicago Park Elementary | 1 | 14.20176315 | 656.5 | | 18.18670163 | | 669.5769412 | -13.07694118 | 171.0063907 | |
| 3 | 3 | Woodside Elementary | 1 | 14.54213619 | 695.2999878 | | | | 668.6118718 | 26.68811596 | 712.2555337 | |
| 4 | 4 | Fort Jones Union Elementary | 1 | 14.70588207 | 666.8499756 | | X(bar) second split | | 668.1475985 | -1.297622893 | 1.683825172 | |
| 5 | 5 | Hillsborough City Elementary | 1 | 15.13898468 | 698.25 | | 21.09414889 | | 666.9196104 | 31.33038964 | 981.5933148 | |
| 6 | 6 | Dunsmuir Elementary | 1 | 15.22435951 | 646.4000244 | | | | 666.6775447 | -20.27752028 | 411.1778289 | |
| 7 | 7 | Loleta Union Elementary | 1 | 15.25885487 | 653.8499756 | | Y(bar) first split | | 666.579739 | -12.72976341 | 162.0468765 | |
| 8 | 8 | Montague Elementary | 1 | 15.27272701 | 635.4500122 | | 658.2783319 | | 666.5404069 | -31.09039474 | 966.6126449 | |
| 9 | 9 | Liberty Elementary | 1 | 15.29304028 | 681.2999878 | | | | 666.4828121 | 14.81717565 | 219.5486943 | |
| 10 | 10 | Portola Valley Elementary | 1 | 15.40704155 | 699.0999756 | | Y(bar) second split | | 666.1595811 | 32.94039449 | 1085.069589 | |
| 11 | 11 | Big Springs Union Elementary | 1 | 15.4285717 | 643.5 | | 650.0347642 | | 666.0985361 | -22.59853606 | 510.6938322 | |
| 12 | 12 | Etna Union Elementary | 1 | 15.51246643 | 655.0500488 | | | | 665.860667 | -10.81061815 | 116.8694648 | |
| 13 | 13 | Latrobe Elementary | 1 | 15.59139729 | 673.9000244 | | B1 | | 665.636872 | 8.263152356 | 68.27968687 | |
| 14 | 14 | Kenwood Elementary | 1 | 15.60418606 | 681.2999878 | | -2.835328397 | | 665.6006117 | 15.69937612 | 246.4704105 | |
| 15 | 15 | Fort Ross Elementary | 1 | 15.88235283 | 678.4000244 | | | | 664.8119175 | 13.58810686 | 184.6366481 | |
| 16 | 16 | Raisin City Elementary | 1 | 16.22857094 | 627.0999756 | | Intercept | | 663.8302755 | -36.7302999 | 1349.114931 | |
| 17 | 17 | Reed Union Elementary | 1 | 16.26228523 | 693.9500122 | | 709.8436035 | | 663.7346844 | 30.21532778 | 912.9660331 | |
| 18 | 18 | Two Rock Union Elementary | 1 | 16.29310226 | 673.0499878 | | | | 663.647308 | 9.402679784 | 88.41038712 | |
| 19 | 19 | West Park Elementary | 1 | 16.31168747 | 635.75 | | | | 663.5946128 | -27.84461284 | 775.3224643 | |

| Obs | ob | district | split | str | testscr | | | | Predicted | Residuals | Squared Residuals | SSR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | McCloud Union Elementary | 1 | 14 | 635.5999756 | | X(bar) first split | | =$H$18+($H$15*E2) | =F2-J2 | =K2^2 | =SUM(L2:L421) |
| 2 | 2 | Chicago Park Elementary | 1 | 14.20176315 | 656.5 | | =AVERAGE(E2:E211) | | =$H$18+($H$15*E3) | =F3-J3 | =K3^2 | |
| 3 | 3 | Woodside Elementary | 1 | 14.54213619 | 695.2999878 | | | | =$H$18+($H$15*E4) | =F4-J4 | =K4^2 | |
| 4 | 4 | Fort Jones Union Elementary | 1 | 14.70588207 | 666.8499756 | | X(bar) second split | | =$H$18+($H$15*E5) | =F5-J5 | =K5^2 | |
| 5 | 5 | Hillsborough City Elementary | 1 | 15.13898468 | 698.25 | | =AVERAGE(E212:E421) | | =$H$18+($H$15*E6) | =F6-J6 | =K6^2 | |
| 6 | 6 | Dunsmuir Elementary | 1 | 15.22435951 | 646.4000244 | | | | =$H$18+($H$15*E7) | =F7-J7 | =K7^2 | |
| 7 | 7 | Loleta Union Elementary | 1 | 15.25885487 | 653.8499756 | | Y(bar) first split | | =$H$18+($H$15*E8) | =F8-J8 | =K8^2 | |
| 8 | 8 | Montague Elementary | 1 | 15.27272701 | 635.4500122 | | =AVERAGE(F1:F211) | | =$H$18+($H$15*E9) | =F9-J9 | =K9^2 | |
| 9 | 9 | Liberty Elementary | 1 | 15.29304028 | 681.2999878 | | | | =$H$18+($H$15*E10) | =F10-J10 | =K10^2 | |
| 10 | 10 | Portola Valley Elementary | 1 | 15.40704155 | 699.0999756 | | Y(bar) second split | | =$H$18+($H$15*E11) | =F11-J11 | =K11^2 | |
| 11 | 11 | Big Springs Union Elementary | 1 | 15.4285717 | 643.5 | | =AVERAGE(F212:F421) | | =$H$18+($H$15*E12) | =F12-J12 | =K12^2 | |
| 12 | 12 | Etna Union Elementary | 1 | 15.51246643 | 655.0500488 | | | | =$H$18+($H$15*E13) | =F13-J13 | =K13^2 | |
| 13 | 13 | Latrobe Elementary | 1 | 15.59139729 | 673.9000244 | | B1 | | =$H$18+($H$15*E14) | =F14-J14 | =K14^2 | |
| 14 | 14 | Kenwood Elementary | 1 | 15.60418606 | 681.2999878 | | =(H12-H9)/(H6-H3) | | =$H$18+($H$15*E15) | =F15-J15 | =K15^2 | |
| 15 | 15 | Fort Ross Elementary | 1 | 15.88235283 | 678.4000244 | | | | =$H$18+($H$15*E16) | =F16-J16 | =K16^2 | |
| 16 | 16 | Raisin City Elementary | 1 | 16.22857094 | 627.0999756 | | Intercept | | =$H$18+($H$15*E17) | =F17-J17 | =K17^2 | |
| 17 | 17 | Reed Union Elementary | 1 | 16.26228523 | 693.9500122 | | =(AVERAGE(F2:F421) -(H15*AVERAGE(E2:E421))) | | =$H$18+($H$15*E18) | =F18-J18 | =K18^2 | |
| 18 | 18 | Two Rock Union Elementary | 1 | 16.29310226 | 673.0499878 | | | | =$H$18+($H$15*E19) | =F19-J19 | =K19^2 | |
| 19 | 19 | West Park Elementary | 1 | 16.31168747 | 635.75 | | | | =$H$18+($H$15*E20) | =F20-J20 | =K20^2 | |

- The sum of squared residuals here is 144,778, while in model 1 it is 144,315. These numbers are very close, but the method of semi-averages produces 463 more values of the residual than the original linear model. This number is expected to be slightly larger than in our OLS model, as OLS is designed to minimize SSR, so any other predictive line (such as the one generated by our method of semi-averages) will have a larger SSR