Regression: Test Score Evaluation

Katelyn Patricio

2025-07-01

Goal: Evaluate how student–teacher ratio and school funding affect academic performance in California schools by modeling their relationship with test scores, and identify which factor has a more significant impact.

```
library(AER)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

```
data("CASchools")

head(CASchools)
```

```
##   district                          school  county grades students teachers
## 1    75119              Sunol Glen Unified Alameda  KK-08      195    10.90
## 2    61499            Manzanita Elementary   Butte  KK-08      240    11.15
## 3    61549     Thermalito Union Elementary   Butte  KK-08     1550    82.90
## 4    61457 Golden Feather Union Elementary   Butte  KK-08      243    14.00
## 5    61523        Palermo Union Elementary   Butte  KK-08     1335    71.50
## 6    62042         Burrel Union Elementary  Fresno  KK-08      137     6.40
##   calworks   lunch computer expenditure    income   english  read  math
## 1   0.5102  2.0408       67    6384.911 22.690001  0.000000 691.6 690.0
## 2  15.4167 47.9167      101    5099.381  9.824000  4.583333 660.5 661.9
## 3  55.0323 76.3226      169    5501.955  8.978000 30.000002 636.3 650.9
## 4  36.4754 77.0492       85    7101.831  8.978000  0.000000 651.9 643.5
## 5  33.1086 78.4270      171    5235.988  9.080333 13.857677 641.8 639.9
## 6  12.3188 86.9565       25    5580.147 10.415000 12.408759 605.7 605.4
```

```r
CASchools$avg_test_scores <- (CASchools$read + CASchools$math) / 2
CASchools$ratio <- CASchools$students / CASchools$teachers
```

Create variables of average test scores and student to teacher ratio for model creation purposes

```r
m <- lm(avg_test_scores ~ ratio + expenditure, data = CASchools)

summary(m)
```

Fit a linear regression model to predict test scores based on student to teacher ratio and school funding

```
##
## Call:
## lm(formula = avg_test_scores ~ ratio + expenditure, data = CASchools)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.507 -14.403   0.407  13.195  48.392
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 675.577176  19.562221  34.535   <2e-16 ***
## ratio        -1.763216   0.610914  -2.886   0.0041 **
## expenditure   0.002487   0.001823   1.364   0.1733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.56 on 417 degrees of freedom
## Multiple R-squared:  0.05545,    Adjusted R-squared:  0.05092
## F-statistic: 12.24 on 2 and 417 DF,  p-value: 6.824e-06
```

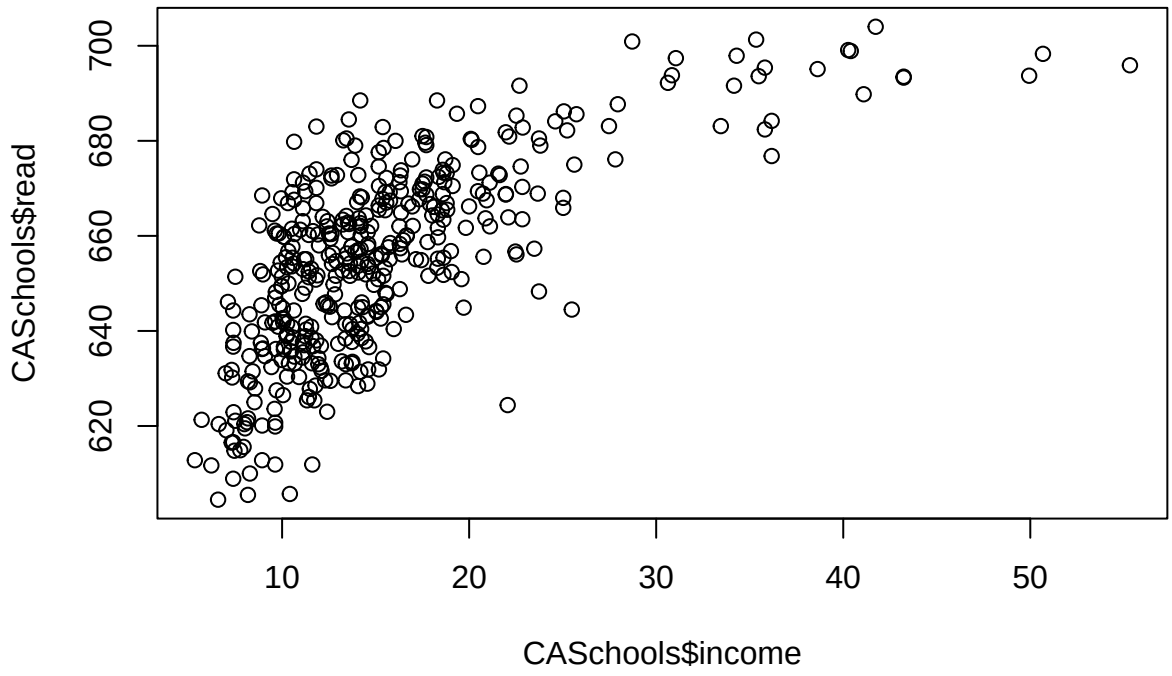
```r
m2 <- lm(read ~ income, data = CASchools)

summary(m2)
```

```
##
## Call:
## lm(formula = read ~ income, data = CASchools)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.665 -10.113   0.998  10.675  35.742
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 625.22768    1.65072  378.76   <2e-16 ***
## income        1.94187    0.09749   19.92   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.42 on 418 degrees of freedom
## Multiple R-squared:  0.487,  Adjusted R-squared:  0.4857
## F-statistic: 396.7 on 1 and 418 DF,  p-value: < 2.2e-16
```

```
plot(CASchools$income, CASchools$read)
```



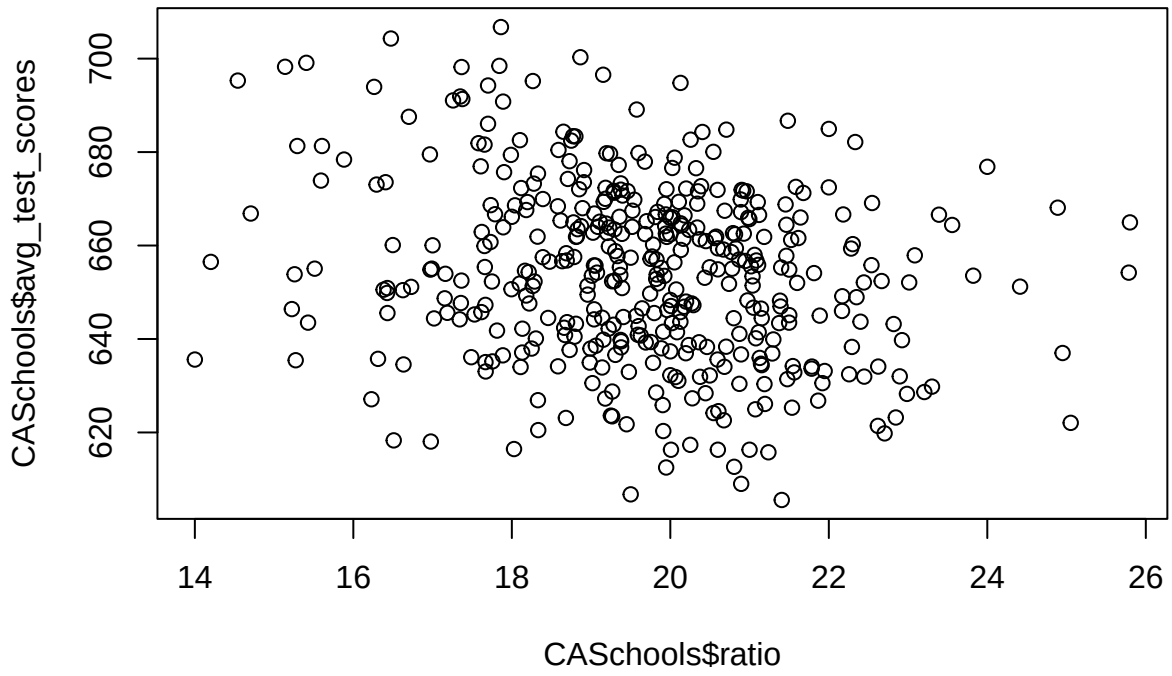

```
# What is the story of this?
m3 <- lm(avg_test_scores ~ income + ratio, data = CASchools)

summary(m3)
```

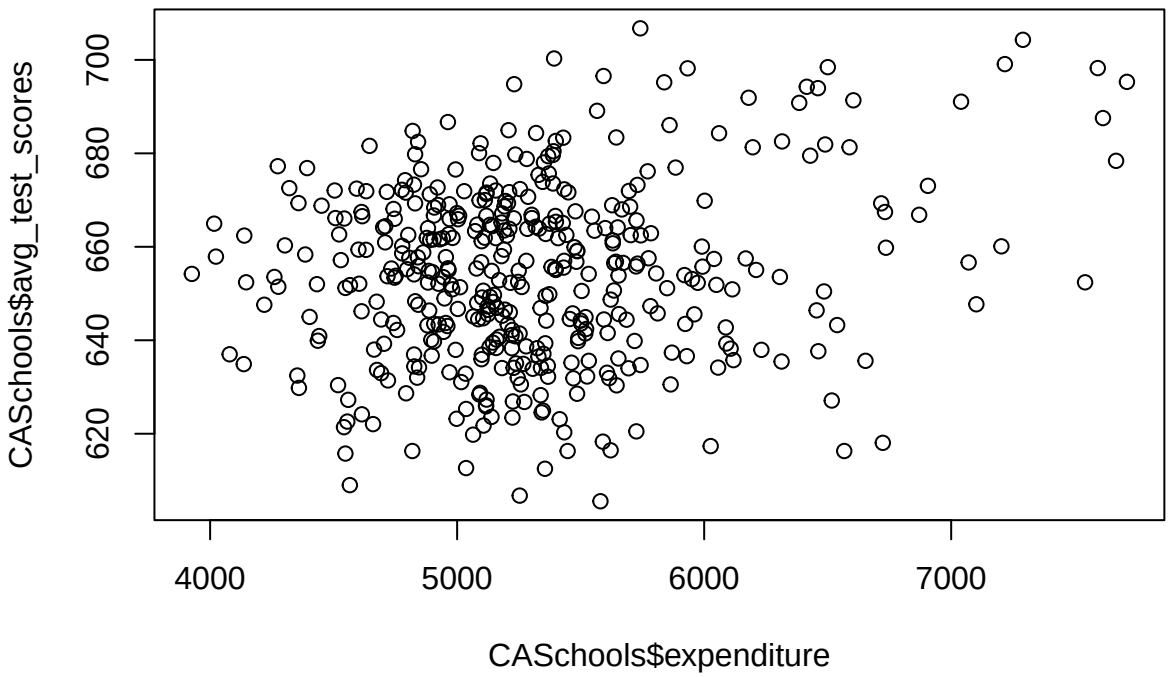
```
##
## Call:
## lm(formula = avg_test_scores ~ income + ratio, data = CASchools)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.608  -9.052   0.707   9.259  31.898
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 638.72916    7.44908  85.746   <2e-16 ***
```

```
## income        1.83911    0.09279  19.821   <2e-16 ***
## ratio        -0.64874    0.35440  -1.831   0.0679 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.35 on 417 degrees of freedom
## Multiple R-squared:  0.5115, Adjusted R-squared:  0.5091
## F-statistic: 218.3 on 2 and 417 DF,  p-value: < 2.2e-16
```

```
# Plot individual variables against test scores to visualize relationships
plot(CASchools$ratio, CASchools$avg_test_scores)
```




```
plot(CASchools$expenditure, CASchools$avg_test_scores)
```



Conclusion:

Student-to-teacher ratio has a significant negative impact on student scores, with each increase in ratio decreasing test scores by approximately 1.76 points. The p-value of 0.0041 with two stars indicates a significant relationship. On the other hand, school funding (expenditure) has a positive but non-significant effect, increasing test scores by about 0.0025 points, with a p-value greater than 0.05, suggesting this relationship is not meaningful. When income is included in the model, it has the largest positive impact on test scores, increasing scores by 1.84 points per unit increase, with a highly significant p-value. The student-to-teacher ratio remains negatively associated but is only marginally significant (p = 0.0679).