An empirical example for simple regression

Q: Are test scores of students affected by student-to-teacher ratios?

The California Test Score Data Set

Test scores and class sizes in 1999 in 420 California school districts that serve kindergarten through 8^{th} grade (n = 420)

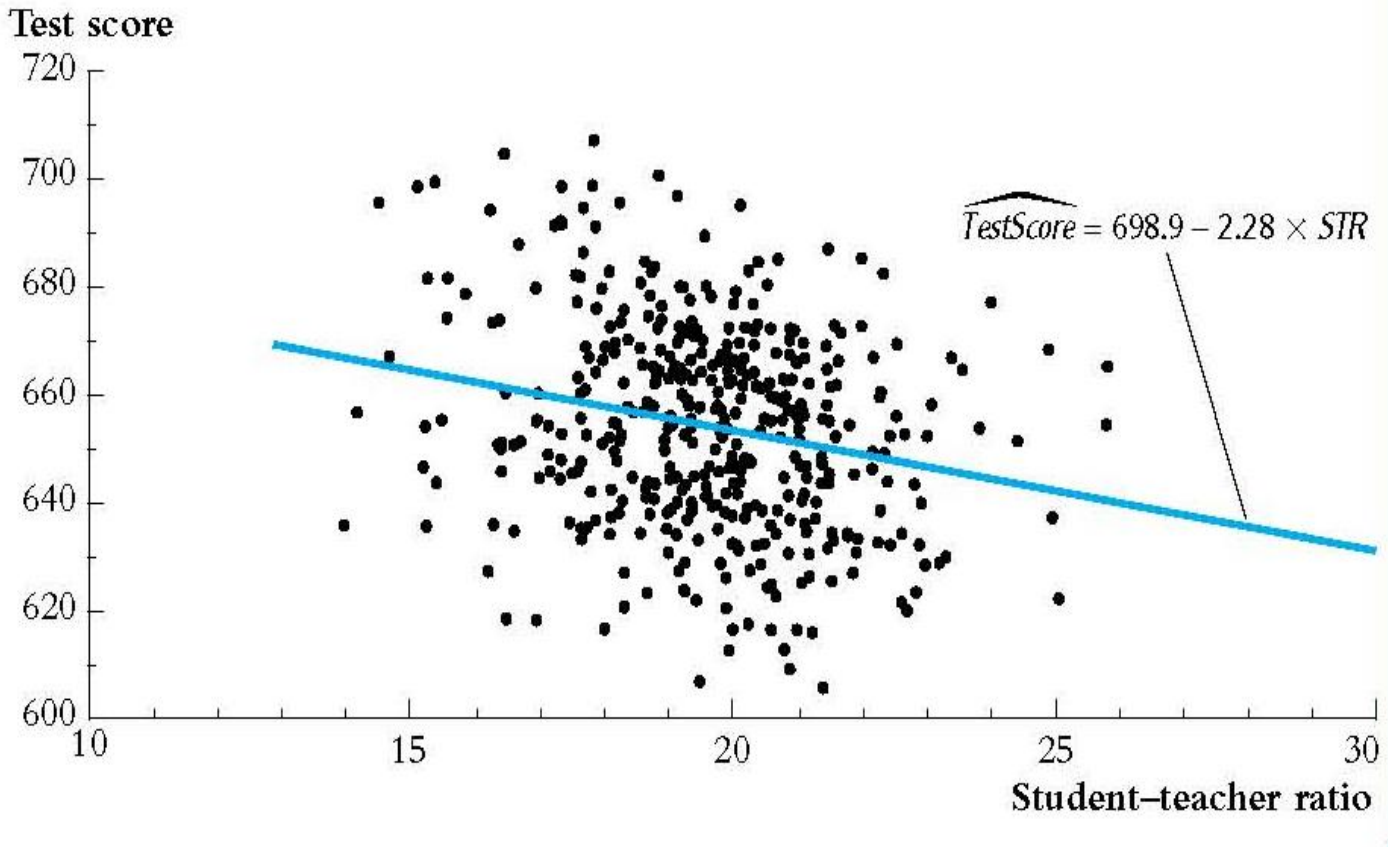
- Dependent variable: 5th grade test scores (Stanford-9 achievement test, combined math and reading), district average

- Indep. variable: Student-teacher ratio (STR) = no. of students in the district divided by no. full-time equivalent teachers

TABLE 4.1 Summary of the Distribution of Student–Teacher Ratios and Fifth-Grade Test Scores for 420 K–8 Districts in California in 1998

			Percentile						
	Average	Standard Deviation	10%	25%	40%	50% (median)	60%	75%	90%
Student–teacher ratio	19.6	1.9	17.3	18.6	19.3	19.7	20.1	20.9	21.9
Test score	665.2	19.1	630.4	640.0	649.1	654.5	659.4	666.7	679.1

This table doesn't tell us anything about the relationship between test scores and the *STR*.



Estimated slope $= \hat{\beta}_1 = -2.28$

Estimated intercept $= \hat{\beta}_0 = 698.9$

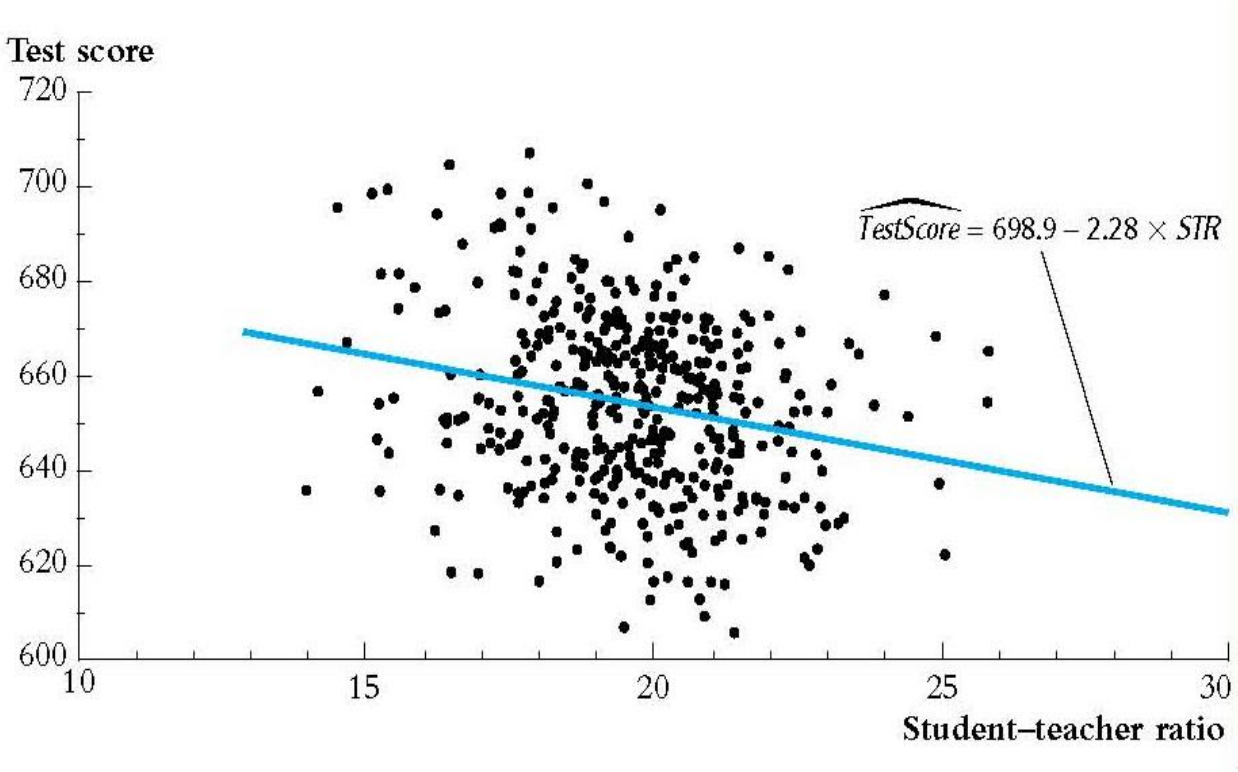
Estimated regression line: $TestScore = 698.9 - 2.28 \times STR$

Interpretation of the estimated slope and intercept

$$TestScore = 698.9 - 2.28 \times STR$$

- Districts with one more student per teacher on average have test scores that are 2.28 points lower.
- That is, $\frac{\Delta \text{Test score}}{\Delta STR} = -2.28$.

Predicted values & residuals:



One of the districts in the data set is Antelope, CA, for which *STR* = 19.33 and *Test Score* = 657.8

predicted value: $\hat{Y}_{Antelope} = 698.9 - 2.28 \times 19.33 = 654.8$

residual: $\hat{u}_{Antelope} = 657.8 - 654.8 = 3.0$

OLS regression: Eviews output

Dependent Variable: TESTSCR
Method: Least Squares
Date: 09/21/15 Time: 12:23
Sample: 1 420
Included observations: 420

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	698.9330	9.467491	73.82451	0.0000
STR	-2.279808	0.479826	-4.751327	0.0000

R-squared	0.051240	Mean dependent var	654.1565
Adjusted R-squared	0.048970	S.D. dependent var	19.05335
S.E. of regression	18.58097	Akaike info criterion	8.686903
Sum squared resid	144315.5	Schwarz criterion	8.706143
Log likelihood	-1822.250	Hannan-Quinn criter.	8.694507
F-statistic	22.57511	Durbin-Watson stat	0.129062
Prob(F-statistic)	0.000003		