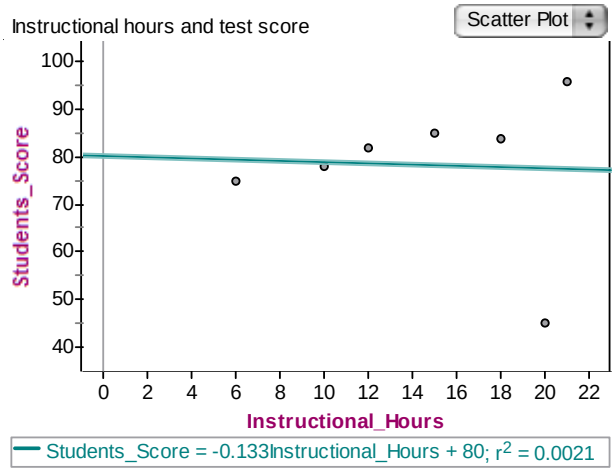
The Effect of Outliers

Example: To evaluate the performance of one of its instructors, a driving school tabulates the number of hours of instruction and the driving-test scores for the instructor's students.

Instructional hours and test score

	Instructional_Hours	Students_Score
1	10	78
2	15	85
3	21	96
4	6	75
5	18	84
6	20	45
7	12	82

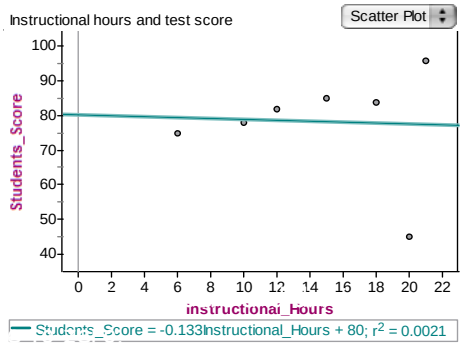


a) What assumptions are being made? (ie. What relationship should be shown if the instructor is an effective instructor) Is this a reasonable assumption?

Answer: The management of the driving school is assuming that the correlation between instructional hours and test scores is an indication of the instructor's teaching skills. Such relationship could be difficult to prove. However, the assumption would be reasonable if the driving school has found that some instructors have consistently strong correlations between the time spent with their students and the student's test scores while other instructors have consistently weaker correlations.

b) Find the correlation coefficient, r, and find the equation of the Line of Best Fit in y = mx + b form)
r=square root(0.0021)=0.0458, so the correlation coefficient is 0.0458.

The equation of the line of Best Fit is
Student_Scores=-.133xInstructional_Hours+80



c) Do these results suggest the instructor is an effective teacher?

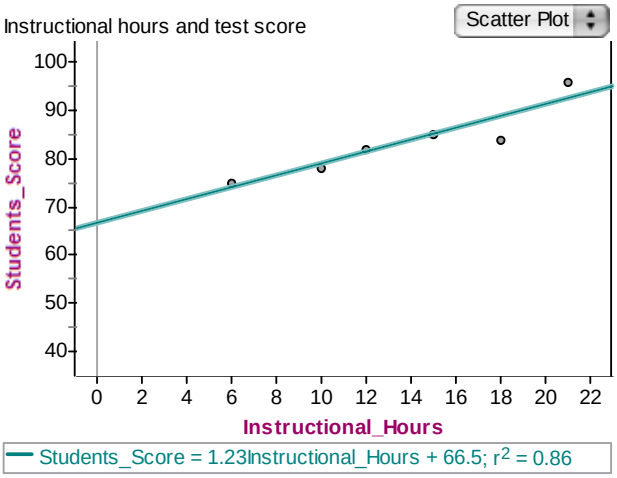
Answer: No, these results do not suggest that the instructor is an effective teacher. The correlation coefficient, r, is 0.0458 suggesting that there is zero linear correlation, and the line of best fit even has a negative slope.

d) Is there any unusual data? (ie. any outliers?)

Answer: (20, 45) is substantially below all the other data points. This point is an outlier.

e) Determine the effects of the outlier on our analysis.

Answer: Removing the outlier from our data and repeating the analysis we obtain the following line of best fit. The correlation coefficient, r, is 0.93, indicating a strong positive linear correlation between the number of instructional hours and the driver's test scores. This result suggests that the instructor may be an effective teacher after all. However, to do a proper evaluation, you would need a larger set of data, more information about the outlier, or, ideally, both.



Homework: Pg. 51 #1-4 Pg. 71 #4, 6