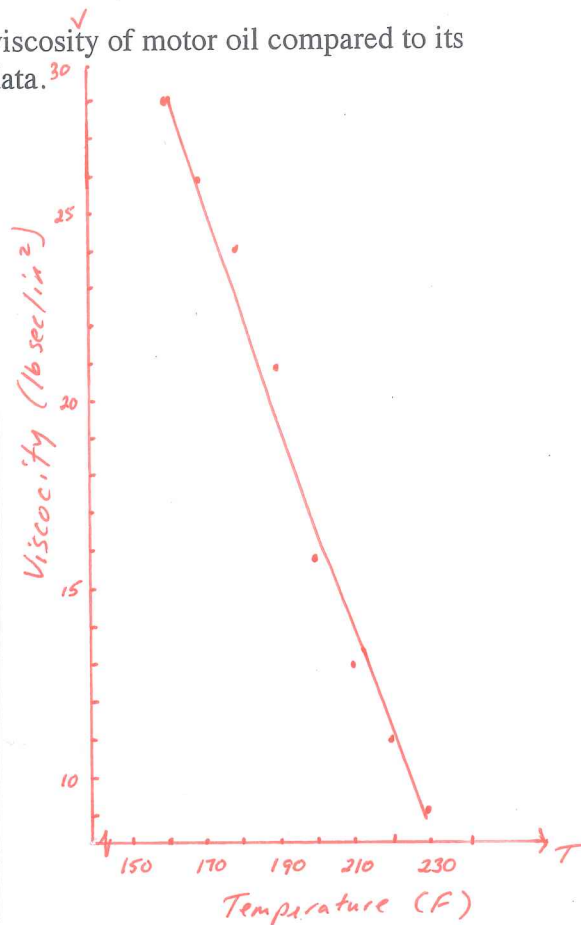


Sec 1.6 Fitting Linear Functions to Data

Laboratory Data: The Viscosity of Motor Oil

a. The following table is given showing the viscosity of motor oil compared to its temperature. Create and label a graph of its data.

T , temperature (°F)	ν , viscosity (lb·sec/in ²)
160	28
170	26
180	24
190	21
200	16
210	13
220	11
230	9



b. Looking at the graph, what ^{does} you notice the scatter plot of data tell you about the viscosity of oil vs temperature (ie explain what its slope means)? *There is a negative relationship between the two. As the temperature increases, the viscosity decreases.*

c. Use a calculator to find the line of best fit for this data set. What does the regression line tell you?

$$\nu = 75.6 - 0.293T$$

The regression line tells us that our data fits it very well, even though no points fall exactly on it.

Interpolation and Extrapolation

(Between known values vs unknown)

Ex. Using the regression line $\nu = 75.6 - 0.293 T$, predict the viscosity of motor oil at 240 F and at 300 F. What do you notice happens? What does this tell you about the graph?

$$\begin{aligned} \nu &= 75.6 - 0.293(240) \\ \nu &= 5.28 \end{aligned}$$

$$\begin{aligned} \nu &= 75.6 - 0.293(300) \\ \nu &= -12.3 \end{aligned}$$

The viscosity becomes negative, which is impossible. The point (300, -12.3) is too far from our data points to be reasonable. The trend observed in the laboratory cannot be extended that far.

Correlation Coefficients

- When a computer or calculator calculates a regression line, it also gives a *correlation coefficient*, r .
- This number lies between -1 and $+1$ and measures how well a particular regression line fits the data.
- If $r = 1$, the data lie exactly on a line of positive slope.
- If $r = -1$, the data lie exactly on a line of negative slope.
- If r is close to 0 , the data may be completely scattered, or there may be a non-linear relationship between the variables.
- For Example 1, $r \approx -0.99$. It is negative because the slope of the line is negative. The fact that r is close to -1 indicates that the regression line is a good fit.

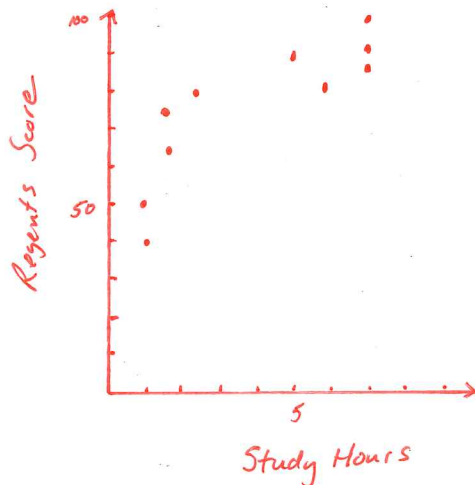
Ex. Find the correlation coefficient for the data set ^{below} above. What does this tell you about the graph and the data?
 $r = .85$

It means it has a fairly strong positive correlation and that most points will be close to the regression line, but not necessarily on it.

Ex. Given the following data set, use your calculator to create a graph (sketch it), find the line of best fit, and ~~find~~ ^{list} the correlation coefficient for the data.

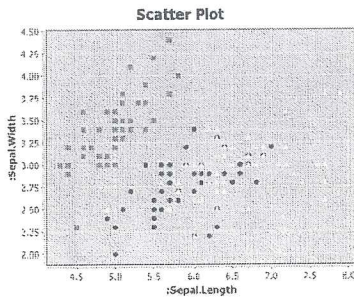
Study Hours	Regents Score
3	80
5	90
2	75
6	80
7	90
1	50
2	65
7	85
1	40
7	100

$$y = 6.27x + 49.78 \quad r = .85$$



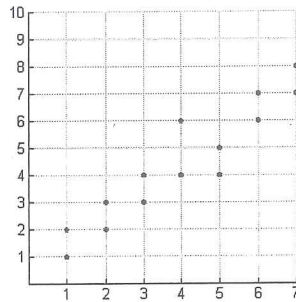
Ex. Given the following data sets, which ^{are} ~~are~~ ^{model} linear functions and what would be a good estimate of a correlation coefficient for each graph?

a.



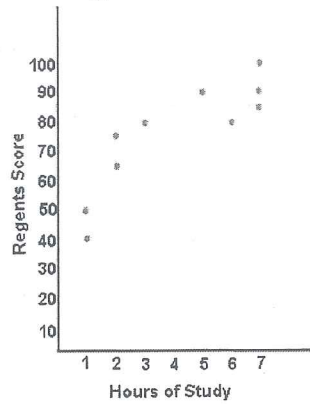
not linear
 $r = .3$

b.



$r = .94$

c.



$r = .85$

The Difference between Relation, Correlation and Causation

- It is important to understand that a high correlation (either positive or negative) between two quantities does *not imply causation*. For example, there is a high correlation between children's reading level and shoe size. However, large feet do not cause a child to read better (or vice versa). Larger feet and improved reading ability are both a consequence of growing older.
- A correlation of $r = 0$ usually implies there is no linear relationship between x and y , but this does not mean there is no relationship at all.

HW: pg 48-50 ~~1-12~~ ALL

1-10