Fitting Linear Functions to Data

Chapter 1

Section 6

Using Real-Life Data

When real data is collected in the laboratory or the field, it is often subject to experimental error. Even if there is an underlying linear relationship between two quantities, real data may not fit this relationship perfectly. However, even if a data set does not perfectly conform to a linear function, we may still be able to use a linear function to help us analyze the data.

The Viscosity of Motor Oil

The viscosity of a liquid, or its resistance to flow, depends on the liquid's temperature. The viscosity of motor oil is a measure of its effectiveness as a lubricant in the engine of a car. Thus, the effect of engine temperature is an important determinant of motor oil performance. The table on the next slide gives the viscosity, v, of motor oil measured in the lab at different temperatures, T.

Motor Oil Viscosity Laboratory Data

Table 1.34 *The measured viscosity, v, of motor oil as a function of the temperature, T*

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

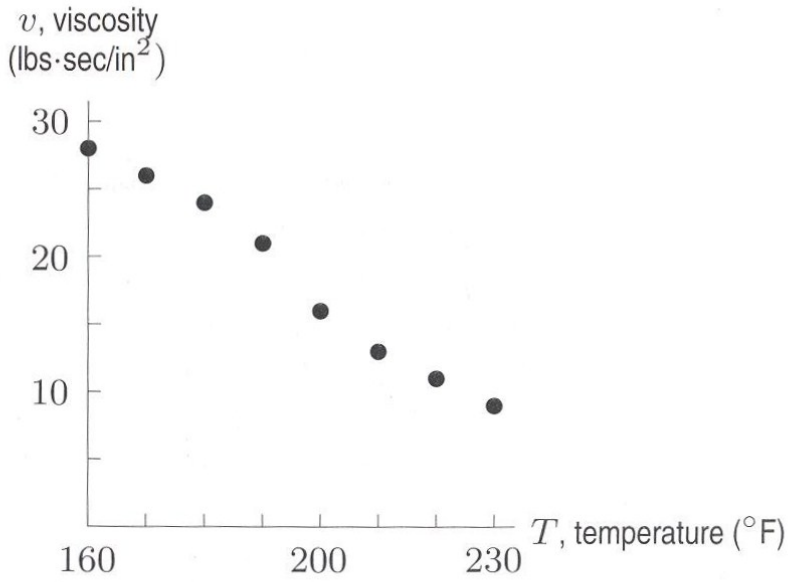
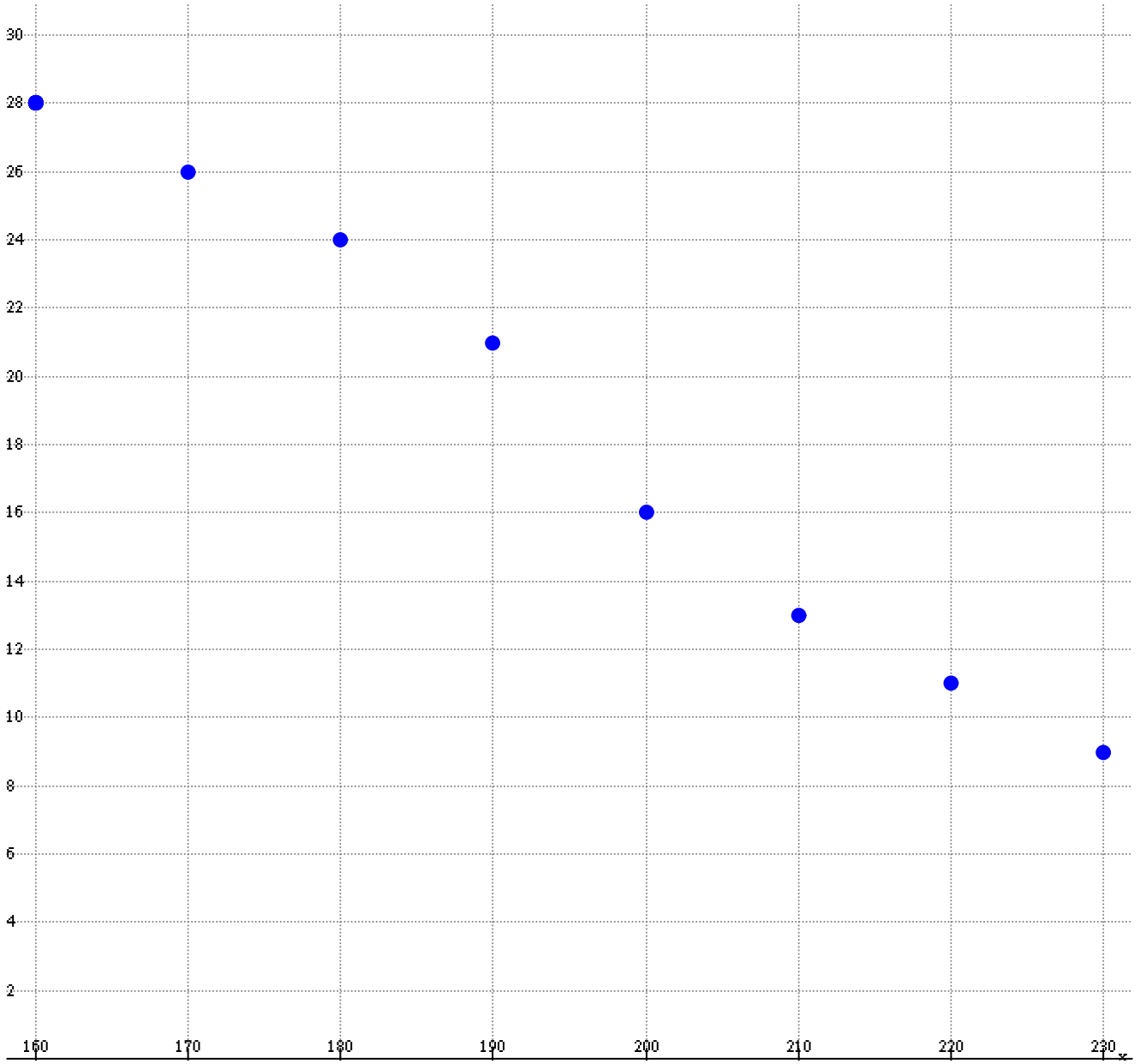
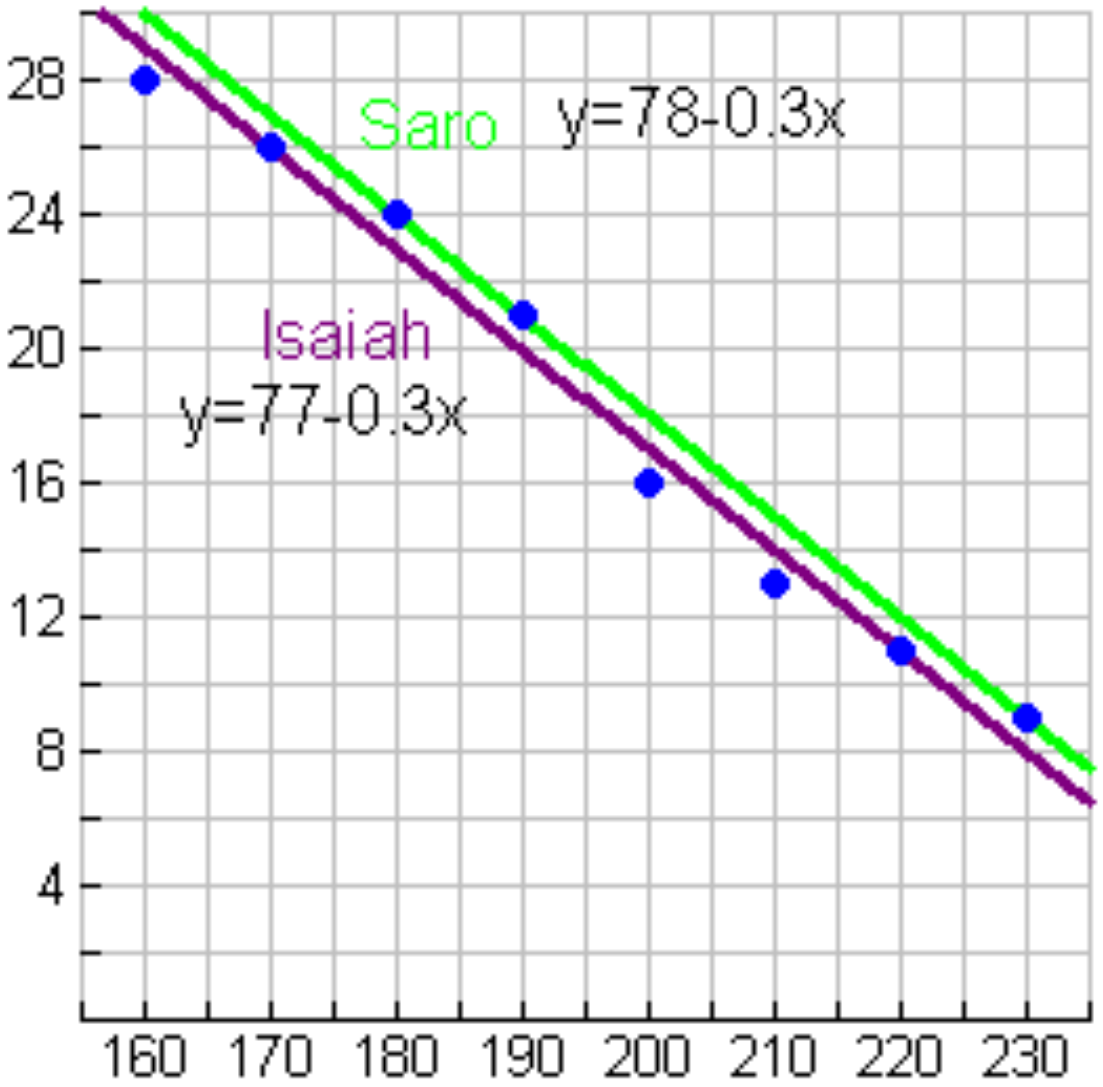


Figure 1.53: The viscosity data from Table 1.34

Enter the temperature data in L1 and the viscosity data in L2 on your calculator and graph.


30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
160
170
180
190
200
210
220
230
x


28
24
20
16
12
8
4
160
170
180
190
200
210
220
230
Saro
y=78-0.3x
Isaiah
y=77-0.3x

How Regression Works

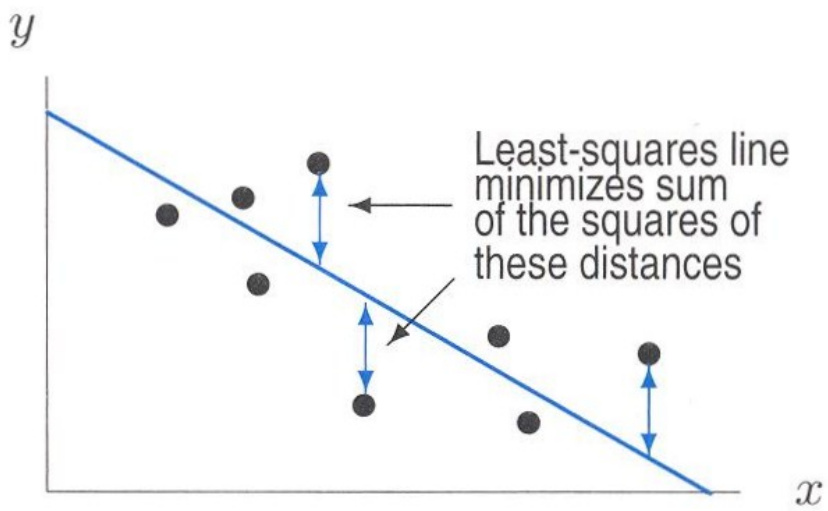


Figure 1.57: A given set of data and the corresponding least-squares regression line

Doing a Manual Fit

Enter x data in L_1 (160, 170, 180, 190, 200, 210, 220, 230), enter y data in L_2 (28, 26, 24, 21, 16, 13, 11, 9).

Turn on Plot1 (scatter plot, L_1, L_2) and Zoomstat

Press STAT, select CALC menu, select Manual-Fit (option D)

Draw best fit line by selecting 2 data points

Best Fit – m = -0.2928, b = 75.607 with a sum of the squares value of 5.785714286

Interpolation

Predict the viscosity of motor oil at 196°F.

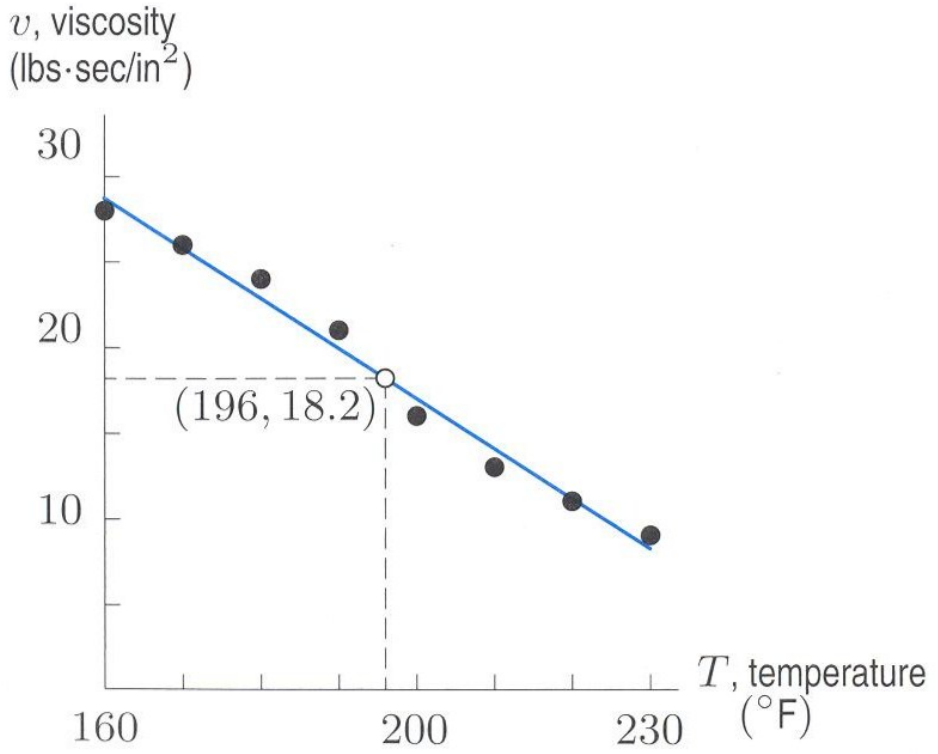


Figure 1.55: Regression line used to predict the viscosity at 196°

Extrapolation

Predict the viscosity of motor oil at 240°F and at 300°F

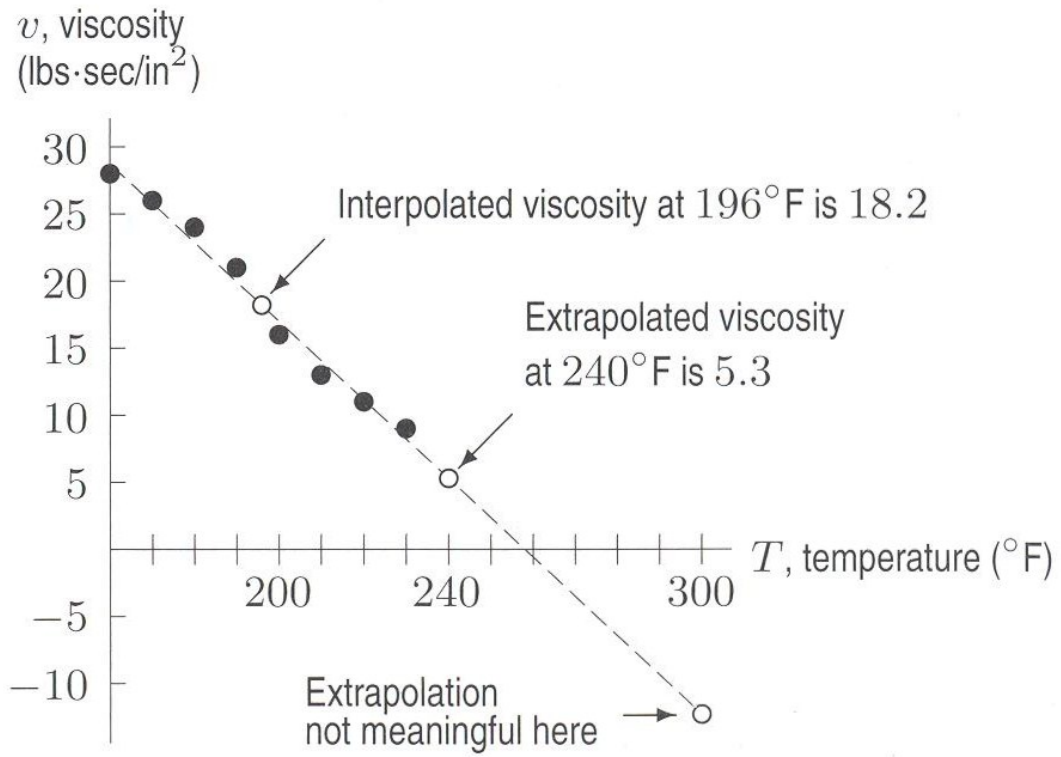


Figure 1.56: The data from Table 1.34 together with the predicted viscosity at $T = 196°$, $T = 240°$, and $T = 300°$

Correlation Coefficient

When a computer or calculator calculates a regression line, it can also give 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

Correlation Examples

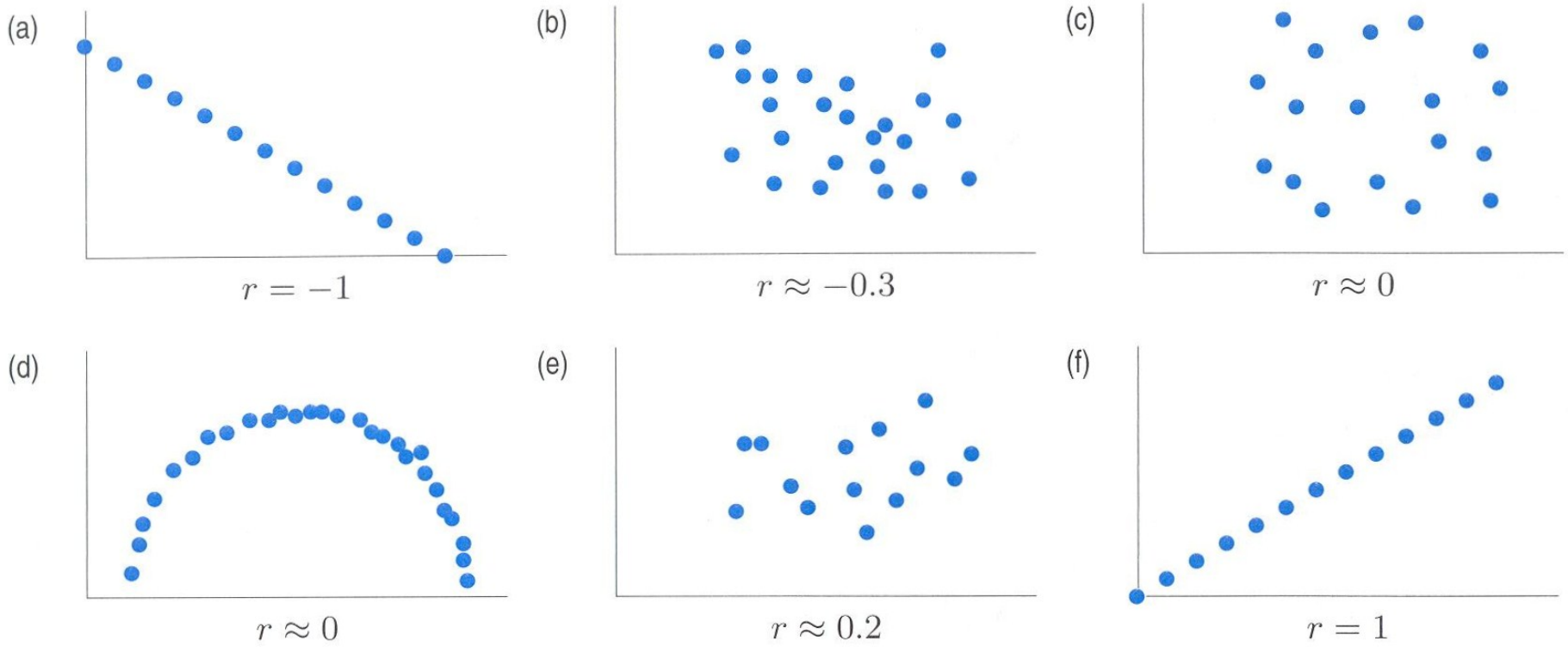


Figure 1.58: Various data sets and correlation coefficients

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 consequence of growing older.

Guessing Correlations

http://cnx.org/content/m11212/latest/

http://serc.carleton.edu/sp/cause/games/examples/13892.html

http://www.rossmanchance.com/applets/guesscorrelation/GuessCorrelation.html

- Not multiple choice