

13.1: Prelude to Linear Regression

Linear regression is a very powerful statistical technique. Many people have some familiarity with regression just from reading the news, where graphs with straight lines are overlaid on scatterplots. Linear models can be used for prediction or to evaluate whether there is a linear relationship between two numerical variables.

Figure 13.1.1 shows two variables whose relationship can be modeled perfectly with a straight line. The equation for the line is

$$y = 5 + 57.49x \quad (13.1.1)$$

Imagine what a perfect linear relationship would mean: you would know the exact value of y just by knowing the value of x . This is unrealistic in almost any natural process. For example, if we took family income x , this value would provide some useful information about how much financial support y a college may offer a prospective student. However, there would still be variability in financial support, even when comparing students whose families have similar financial backgrounds.

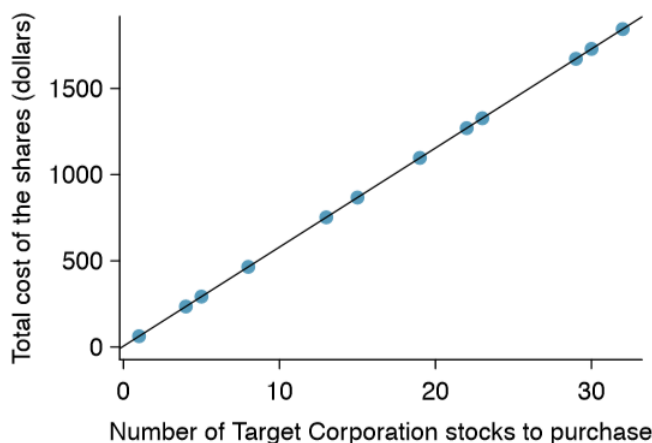


Figure 13.1.1: Requests from twelve separate buyers were simultaneously placed with a trading company to purchase Target Corporation stock (ticker TGT, April 26th, 2012), and the total cost of the shares were reported. Because the cost is computed using a linear formula, the linear fit is perfect.

Linear regression assumes that the relationship between two variables, x and y , can be modeled by a straight line:

$$y = \beta_0 + \beta_1 x \quad (13.1.2)$$

where β_0 and β_1 represent two model parameters (β is the Greek letter beta). These parameters are estimated using data, and we write their point estimates as $\hat{\beta}_0$ and $\hat{\beta}_1$. When we use x to predict y , we usually call x the *explanatory or predictor variable*, and we call y the *response*.

It is rare for all of the data to fall on a straight line, as seen in the three scatterplots in Figure 13.1.2. In each case, the data fall around a straight line, even if none of the observations fall exactly on the line. The first plot shows a relatively strong downward linear trend, where the remaining variability in the data around the line is minor relative to the strength of the relationship between x and y . The second plot shows an upward trend that, while evident, is not as strong as the first. The last plot shows a very weak downward trend in the data, so slight we can hardly notice it. In each of these examples, we will have some uncertainty regarding our estimates of the model parameters, β_0 and β_1 . For instance, we might wonder, should we move the line up or down a little, or should we tilt it more or less?

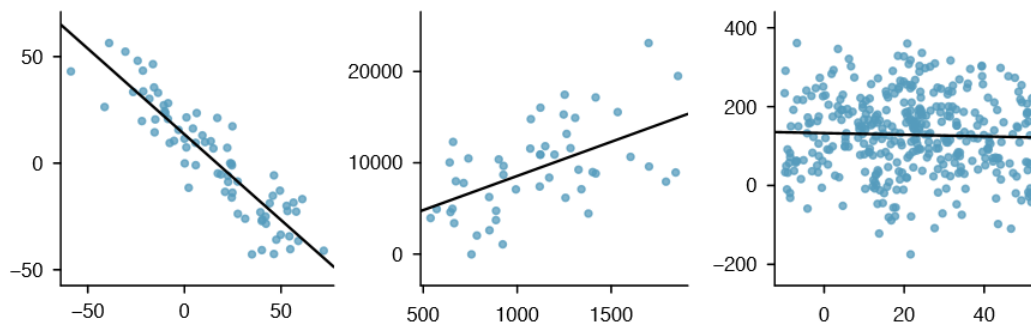


Figure 13.1.1: Three data sets where a linear model may be useful even though the data do not all fall exactly on the line.

As we move forward in this chapter, we will learn different criteria for line-fitting, and we will also learn about the uncertainty associated with estimates of model parameters.

This page titled [13.1: Prelude to Linear Regression](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [David Diez, Christopher Barr, & Mine Çetinkaya-Rundel](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **7.1: Prelude to Linear Regression** by David Diez, Christopher Barr, & Mine Çetinkaya-Rundel is licensed [CC BY-SA 3.0](#). Original source: <https://www.openintro.org/book/os>.