Objective

Find the equation of a regression line.

Notation for the Equation of a Regression Line

	Population Parameter	Sample Statistic
y-intercept of regression equation	β_0	b_0
Slope of regression equation	β_1	b_1
Equation of the regression line	$y = \beta_0 + \beta_1$	$\hat{y} = b_0 + b_1x$

Requirements

1. The sample of paired (x, y) data is a *random* sample of quantitative data.
2. Visual examination of the scatterplot shows that the points approximate a straight-line pattern.
3. Outliers can have a strong effect on the regression equation, so remove any outliers if they are known to be errors. Consider the effects of any outliers that are not known errors.

Note: Requirements 2 and 3 above are simplified attempts at checking these formal requirements for regression analysis:

- For each fixed value of x, the corresponding values of y have a normal distribution.
- For the different fixed values of x, the distributions of the corresponding y-values all have the same standard deviation. (This is violated if part of the scatterplot shows points very close to the regression line while another portion of the scatterplot shows points that are much farther away from the regression line. See the discussion of residual plots in Part 2 of this section.)
- For the different fixed values of x, the distributions of the corresponding y-values have means that lie along the same straight line.

The methods of this section are not seriously affected if departures from normal distributions and equal standard deviations are not too extreme.

Formulas for Finding the Slope b_1 and y-Intercept b_0 in the Regression Equation $\hat{y} = b_0 + b_1x$

Formula 10-3 **Slope:** $b_1 = r\dfrac{s_y}{s_x}$ where r is the linear correlation coefficient, s_y is the standard deviation of the y values, and s_x is the standard deviation of the x values.

Formula 10-4 ***y*-intercept:** $b_0 = \bar{y} - b_1\bar{x}$

The slope b_1 and y-intercept b_0 can also be found using the following formulas that are useful for manual calculations or writing computer programs:

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \qquad b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Rounding the Slope b_1 and the y-Intercept b_0

Round b_1 and b_0 to three significant digits. It's difficult to provide a simple universal rule for rounding values of b_1 and b_0, but this rule will work for most situations in this book. (Depending on how you round, this book's answers to examples and exercises may be slightly different from your answers.)