**Linear regression:** The idea of trying to fit a line as closely as possible to as many points as possible is known as linear regression. The most common technique is to try to fit a line that minimizes the squared distance to each of those points. This is called OLS or Ordinary Least Squares Regression:

\[
\hat{y} = b_0 + b_1 * x_1
\]

**Regression**

One way to make predictions is by using the *regression equation*.

Since our regression estimates form a straight line, we can describe them using an equation in slope-intercept form:

-坡度 of the regression line = \( r \cdot \frac{SD \ of \ y}{SD \ of \ x} \)
-截距 of the regression line = average of \( y \) – slope \( \cdot \) average of \( x \)

**Interpretations**

**SLOPE** = The average increase in \( Y \) associated with a 1-unit increase in \( X \).

**Y-INTERCEPT** = The predicted value of \( Y \) when \( X \) is equal to 0.
Puzzle #1:
Let’s go ahead and find the slope and y-intercept of the regression equation using these 5 summary statistics (rounded):

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carat</td>
<td>0.80</td>
<td>0.50</td>
</tr>
<tr>
<td>Price</td>
<td>3933</td>
<td>3989</td>
</tr>
</tbody>
</table>

Slope (b1):

y-intercept (b0):

Regression Equation:

Use the regression equation on the previous page to predict the price of a diamond that is 2.5 carats.

How do you interpret the slope and the y-intercept in the above equation?

Warning About Regression:
- Predicting y at values of x beyond the range of x in the data is called **extrapolation**.
- This is risky because we have no evidence to believe that the association between x and y remains linear for unseen values of x.
- Extrapolated predictions can be absolutely wrong.