**Regression or Best-fit line**

It describes the approximate rule that the data should follow.

Example:

Diameter (cm)

Age of Tree

After drawing a rectangle around the scatter plot (not including outliers), we will draw a straight line that best fits the direction of the points:

Diameter (cm)

Age of Tree

The line should pass through the middle of the rectangle, parallel to its sides, and as close to as many points as possible.

Next, determine the equation of the line by choosing two points that the line passes through. (Could be part of data or read from the graph)

**y = ax + b**

RECALL: to find the equation of a line passing through two points…

P1 (x1, y1) and P2 (x2 , y2)

**a** = **b** = y1 – a(x1)



What is the purpose of this line?

**Predicting** a value that isn’t part of the data, according to the best-fit line…

* Extend the best-fit line and read from the graph

OR

* Use the equation of the line and plug in the value given for x or for y

**Example**

