

# Linear Regression

1. Plot a **scatter diagram** using the following data points: (1, 5), (2, 3), (3, 4), (4, 2), (5, 2), (6, 1).

Note that the points on the scatter diagram have a linear character. Our goal is to find a line which “fits” the data better than any other line. This line of best fit is called the **least squares regression line**.

The slope  $m$  and the  $y$ -intercept  $b$  of the least squares regression line are found by using the following formulas.

$$m = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$b = \frac{\sum y - m \sum x}{n}$$

Here  $n$  is the number of data points,  $\sum x$  is the sum of the  $x$ -coordinates of the data points,  $\sum y$  is the sum of the  $y$ -coordinates of the data points,  $\sum x^2$  is the sum of the squares of the  $x$ -coordinates of the data points, and  $\sum xy$  is the sum of the products of the  $x$ -coordinate and the  $y$ -coordinate of each data point.

The equation of the least squares regression line is

$$y = mx + b$$

where  $m$  and  $b$  are the solutions to the system of equations which appears above.

2. Find the least squares regression line for the data.

$$n = 6, \sum x = 21, \sum y = 17, \sum x^2 = 91, \sum xy = 47$$

$$m = \frac{6 \cdot 47 - 21 \cdot 17}{6 \cdot 91 - 21^2} = -0.7$$

$$b = \frac{17 - (-0.7) \cdot 21}{6} = 5.3$$

So the least squares regression line is:

$$y = -0.7x + 5.3$$

3. Find the least squares regression line for the data using the linear regression capabilities of your calculator.
4. Graph the least squares regression line on the scatter plot (scatter diagram).
5. Estimate  $y$  when  $x = 7$ .
6. Complete each part using the data given in the table below. Let  $x$  be the number of years since 1992 and  $y$  the net imports of oil in thousands of barrels in country **Z**.

Year	Net imports of oil (thousands of barrels)
1992	7,161
1994	6,938
1996	8,054
1998	8,498
2000	9,764
2002	10,419
2004	10,546
2006	12,097
2007	12,353

- A. Find the least squares regression line. The coefficients of the regression line should be given the the nearest integer.
- B. Plot the scatter diagram and the least squares regression line.
- C. Predict the net imports of oil in barrels in the year 2010.
- D. Predict the year in which the net imports of oil will surpass 14,000,000 barrels.

