

1.5: The Method of Least Squares

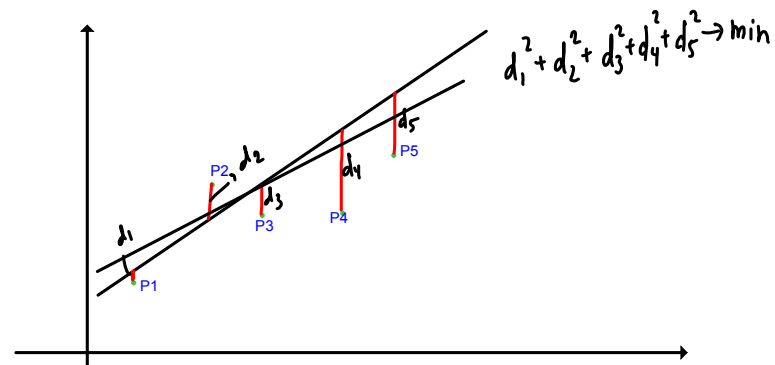
The Method of Least Squares is a general method of finding the "best" fitting formula for a data set.

A **scatter diagram** is the graph of data over a wide range. If the data points seem to be scattered about a straight line, we use the least squares method to find the best fit line, also called the **regression line**.

Suppose, for simplicity, we are given 5 data points:

$$P_1(x_1, y_1), P_2(x_2, y_2), P_3(x_3, y_3), P_4(x_4, y_4), P_5(x_5, y_5)$$

describing the relationship between the two variables x and y . By plotting these data points we obtain a scatter diagram:



The least squares method requires that the sum of the squares of the errors in y-coordinate be a minimum.

To find the linear regression line (- aka - least squares line - aka - line of best fit) we shall use the TI calculator (we skip the mathematical formula).

THE CALCULATOR PROCEDURES FOR LINEAR REGRESSION:

- Enter your x and y values into lists:
 - hit **STAT** and select **1:Edit** . If you have anything in L1 and L2, cursor up to the name of the list, hit **CLEAR** and ENTER.
 - Use the arrow keys and the **ENTER** key to move around while entering your data. Make sure that x under L1 is paired with the corresponding y under L2.
 - **2nd** **Quit** to go back to HOME SCREEN.

- Graph the data points.
 - **2nd** **Y=** **ENTER** (select Plot1).
 - select **On** on the 2nd line, 1st symbol for **Type**, L1 for **Xlist**, L2 for **Ylist**, and your choice of mark for the points. Once this is set up you can turn it on and off from the top of the Y= screen. This setup needs to be done only once(unless your calculator is reset).
 - Hit **ZOOM** and select **9: ZoomStat** to graph the scatter plot. This will have the calculator choose the window settings.

- Find the best fit line.
 - **STAT** **▷** (move to **CALC**) (select **4:LinReg(ax+b)**) **2nd** L1 , **2nd** L2 **ENTER** .
The answer appears on the screen; a is the slope, b is the y intercept
Now the calculator needs to know which lists contain the data. The calculator always assumes that the x -values are in L1 and the y -values are in L2. If this is the case, then press enter and the calculator will compute the regression equation. If your data is in any other list, then you will need to specify the lists. The list can be indicated by pressing **2nd** and any of the numbers 1 to 6.

- Graph the line (You need to paste the function into the Y= table before graphing.).
 - **Y=** **CLEAR** (clear out the old function) **VARs** (select **5:Statistics**) **▷** **▷** (move to EQ, and select **RegEQ**) **ENTER** .
 - hit **GRAPH**

NOTE: Make sure that you turn off **STAT PLOTS** when you are done.

EXAMPLE 1. The following table gives the total sales of drugs (in billions of dollars) in the United States from 1999 ($t = 0$) through 2003:

Year t	0	1	2	3	4
Sales, y	126	144	171	191	216

1999 2000 2001 2002 2003 2004 2005

(a) Look at a scatter plot and decide if a line is a good model for the data. YES

(b) Find an equation of the least-squares line (regression line) for these data.

$$y = ax + b$$

$$a = 22.7$$

$$b = 124.2$$

$$y = 22.7t + 124.2$$

(c) Predict the total sales of drugs in 2005, assuming that the trend is continued.

$t = 6$ for year 2005

$$y = 22.7 \cdot 6 + 124.2 = 260.4$$

\$260.4 billions

(d) In what year would the total sales of drugs in U.S. be 351 billions of dollars?

$$y = 351$$

$$\underbrace{351}_{y_1} = \underbrace{22.7t + 124.2}_{y_2}$$

$$22.7t = 351 - 124.2$$

$$t = \frac{351 - 124.2}{22.7} \approx 10$$

⇓
year 2009