6.3 – Modelling Data with a Line of Best Fit

Date: Apr. 17

Definitions:

- **Line of best fit:** A straight line that best approximates the trend in a scatter plot.
- **Regression function:** A line or curve of best fit, developed through a statistical analysis of data.
- **Interpolation:** The process used to estimate a value within the domain of a set of data, based on a trend.
- **Extrapolation:** The process used to estimate a value outside the domain of a set of data, based on a trend.

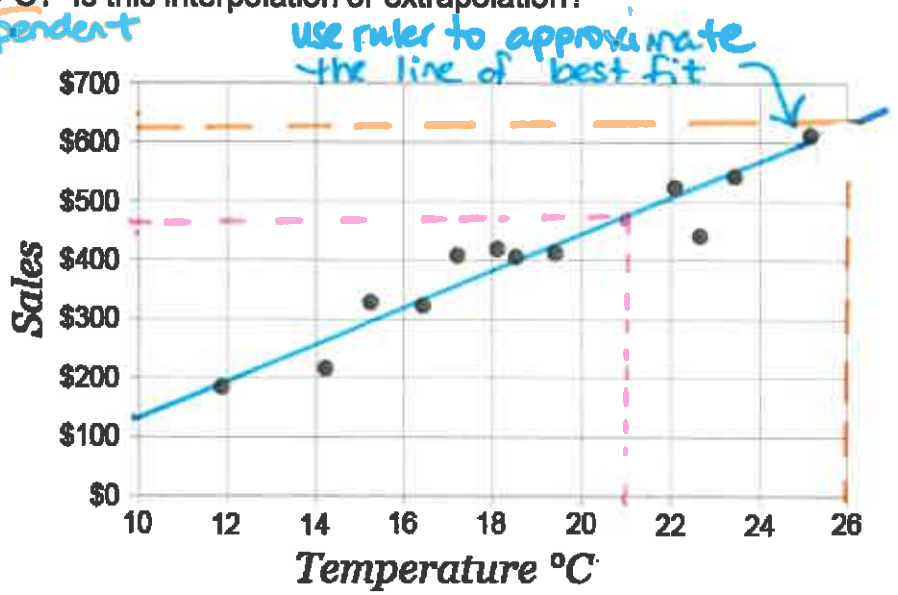
Example: For the data below, determine a line of best fit.

Use the line of best fit to determine expected sales when the temperature is 21°C. Is this interpolation or extrapolation?

What are the expected sales at 26°C? Is this interpolation or extrapolation?

independent → x y ← dependent

Ice Cream Sales vs Temperature	
Temperature °C	**Ice Cream Sales**
14.2°	$215
16.4°	$325
11.9°	$185
15.2°	$332
18.5°	$406
22.1°	$522
19.4°	$412
25.1°	$614
23.4°	$544
18.1°	$421
22.6°	$445
17.2°	$408



@ 21°C, expected sales ≈ $460. (interpolation)

@ 26°C, expected sales ≈ $630 (extrapolation)

Using your Graphing Calculator to Plot Scatter Plots

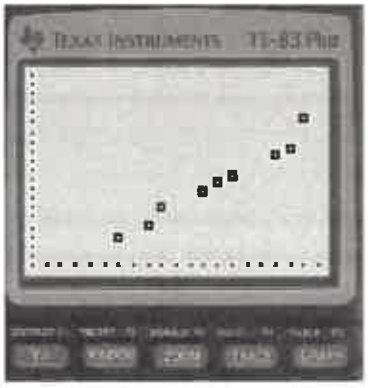
Example: The one-hour record is the farthest distance travelled by bicycle in 1 hour. The table below shows the world-record distances and their dates. Create a line of best fit, and interpolate the results for 2006.

x
y

Year	1996	1998	1999	2002	2003	2004	2007	2008	2009
Distance (km)	78.04	79.14	81.16	82.60	83.72	84.22	86.77	87.12	90.60

On your calculator, press Stat, then select 1:Edit to view the lists. Enter the year into L1 and the distance into L2. Then, press 2nd (second function) and Y= to open up the STAT PLOT menu. Press Enter and turn Plot 1 to On. *

Press graph. The calculator will now plot the data on the graph; however, the window is unlikely set correctly. Press window, and set your window dimensions to fit the data you have just plotted. Press graph.



* turn off if you are no longer using stat plots (will get ERR: INVALID DIMENSION)

Finding the Equation of Best Fit

Estimate the function of the data. The above data appears to be linear, so we will use a line to approximate our data. This is the form y = ax + b.

Press STAT, and move the cursor over to CALC. Select option 4:LinReg(ax+b), and the calculator will return to the main screen showing LinReg(ax+b). Press Enter.

The calculator gives the values for a and b, so you can write your equation as:

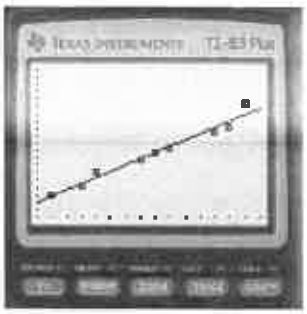
$$y = 0.8585x - 1635.7328$$

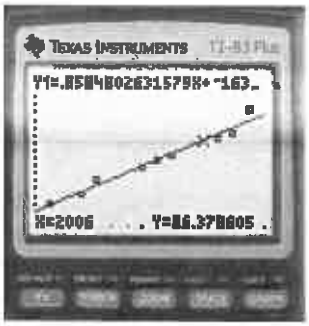
Graphing the Line of Best Fit

Press Y= to bring up the screen for functions.Press VARS and select 5:Statistics. Move over to the EQ heading and select 1:RegEQ.

The calculator will enter the Linear Regression equation you just solved. Press Graph.

You should now have your data points and the line of best fit graphed, as shown below. Use 2nd and CALC to solve for a value, or use the TRACE button to trace the line of best fit to find a point on the line.





to clear lists: STAT 4:ClrList 2nd 1 ENTER ← L1
2nd 2 ENTER ↖ L2

(or) highlight list title and hit CLEAR

Example: Matt buys T-shirts for a company that prints art on T-shirts and then resells them. When buying the T-shirts, the price Matt must pay is related to the size of the order. Five of Matt's past orders are listed in the table below.

$L_1 = x$ $L_2 = y$

Number of Shirts	Cost per Shirt ($)
500	3.25
700	1.95
200	5.20
460	3.51
740	1.69

Matt has misplaced the information from his supplier about price discounts on bulk orders. He would like to get the price-per-shirt below $1.50 on his next order.

a) Use technology to create a scatter plot and determine an equation for the linear regression function that models the data.

$a = -0.0065$
$b = 6.5$

$f(x) =$
or
$y = -0.0065x + 6.5$

* don't forget to clear y= and reset window!

b) What do the slope and y-intercept of the equation of the linear regression function represent in this context?

y-int: (6.5) base price per order / fixed cost per shirt.

slope: rate at which the cost per shirt decreases as each additional shirt is ordered.

c) Use the linear regression function to extrapolate the size of order necessary to achieve the price of $1.50 per shirt.

what value of x (number of shirts) gives us $y = 1.5$? (y: cost per shirt.)

plot $Y = 1.5$
find intersection $\Rightarrow x = 769.2$
2nd TRACE
5: intersect

770 shirts