# Math 30-2: U5L3 Teacher Notes

## Modelling Data with a Line of Best Fit

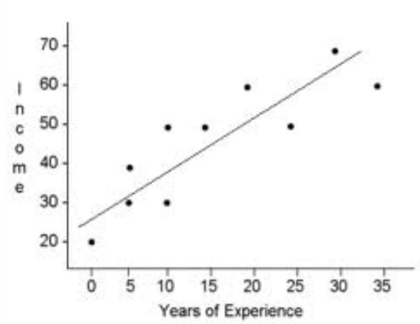
## Key Math Learnings:

**By the end of this lesson, you will learn the following concepts:**

- Graph data and determine the polynomial function that best approximates the data.
- Interpret the graph of a polynomial function that models a situation, and explain the reasoning.
- Solve, using technology, a contextual problem that involves data that is best represent by graphs of polynomial functions, and explain the reasoning.

## Scatterplots

**Scatterplot:** A graph that displays the association between two pieces of data.



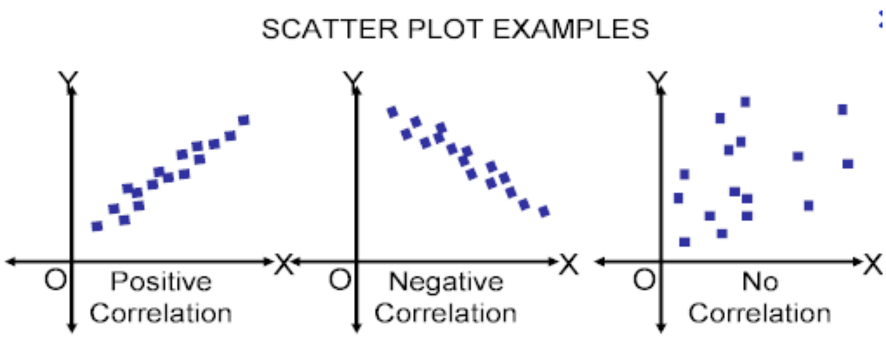

This scatterplot shows the association between the Years of Experience and Income. The **Independent Variable** (years of income) is plotted on the horizontal axis and the **Dependent Variable** (income) is plotted on the vertical axis.

## Things to consider:

If the points cluster in a band running from lower left to upper right, there is a **positive correlation** (if x increases, y increases).

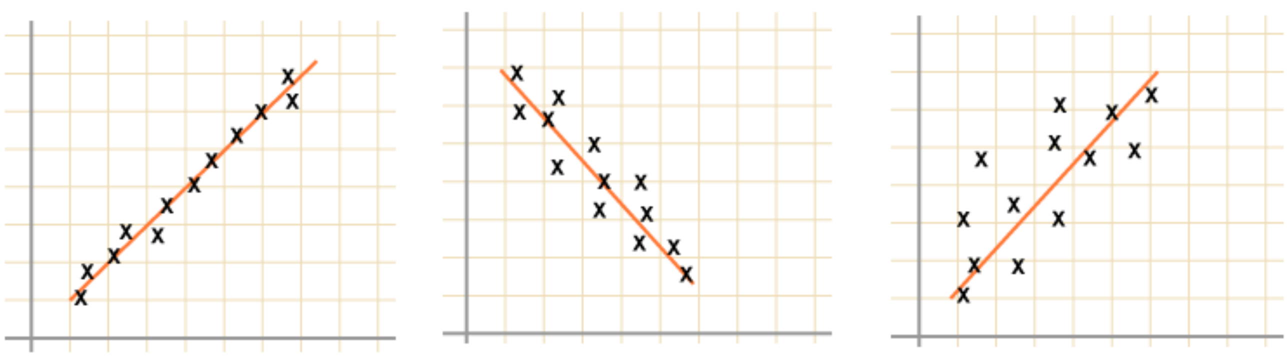
If the points cluster in a band from upper left to lower right, there is a **negative correlation** (if x increases, ydecreases).

If it is hard to see where you would draw a line, and if the points show no significant clustering, there is probably **no correlation.**



## Line of Best Fit

If the points on a scatterplot follow a **linear trend**, technology can be used to determine and graph the equation of the **line of best fit**. The **line of best fit** is drawn so that the points are evenly distributed on either side of the line. It is considered to be the line or curve (next lesson) that best represents the data plotted.



A **line of best fit** is drawn through a scatterplot to find the direction of an association between two variables. This line of best fit can then be used to make predictions.

## Regression Function

**Regression Function:** A line or curve of best fit, developed through a statistical analysis of the data. We find the regression function use technology.

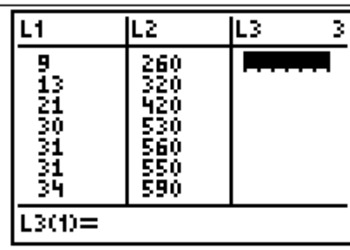
**How to Use the Graphing Calculator to find the Linear Regression Function**

<table>
<tr>
<td>

**1.** Enter the data in the calculator lists. Place the data in $L_1$ and $L_2$.

**STAT, #1Edit,** type values into the lists

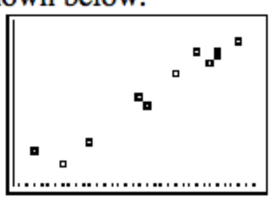
</td>
<td>



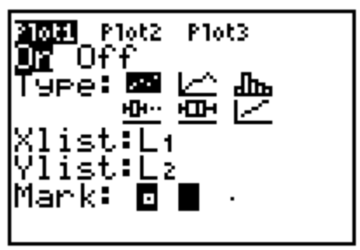

</td>
</tr>
<tr>
<td>

**2.** Prepare a scatter plot of the data. Set up for the scatterplot.

$2^{nd}$ **StatPlot** - choose the first icon - choices shown at right. Choose **ZOOM #9 ZoomStat**. Graph shown below.



</td>
<td>



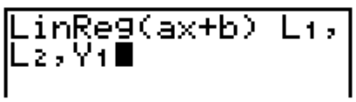

</td>
</tr>
</table>

**3.** Have the calculator determine the line of best fit.

**STAT → CALC #4 LinReg(ax+b)**

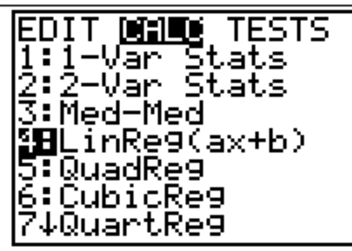
Include the parameters **$L_1$, $L_2$, $Y_1$.**

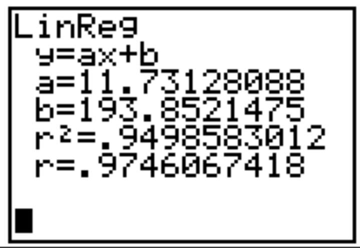
($Y_1$ comes from **VARS → YVARS, #Function, $Y_1$**)




You now have the values of a and b needed to write the equation of the actual **line of best fit.** See values at the right.

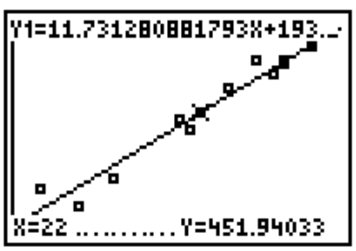
$y = 11.73128088x + 193.8521475$

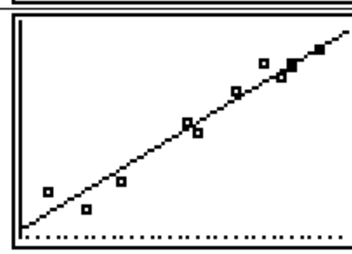



**4.** Graph the line of best fit. Simply hit **GRAPH**.

To get a predicted value **within the window**, hit TRACE, up arrow, and type the desired value. The screen below shows x = 22.



Click the icon to watch a video on how to use your graphing calculator to find the regression function.

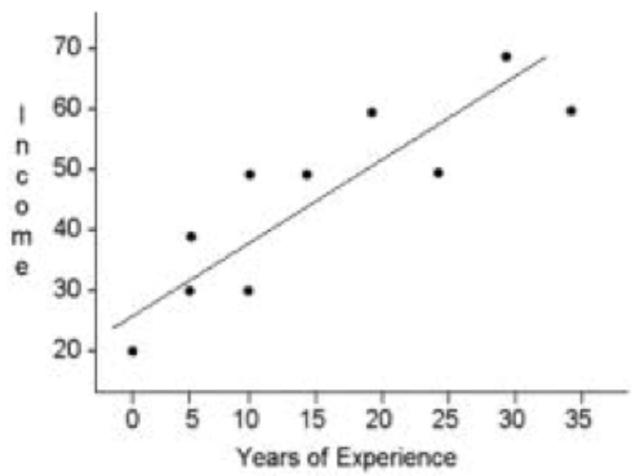
## Interpolation and Extrapolation

Like stated above, we use the line of best fit and the linear regression function to make predictions of the data.

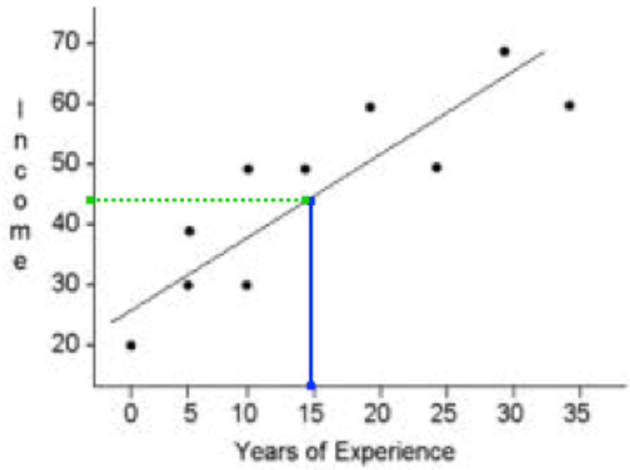
When we use the graph itself to make prediction then we have used **Interpolation.** We have estimated a value **INSIDE** the domain of a set of data.

**Example:**

Use the graph below to predict what your income will be after 15 years of experience.



***Solution:***




Since 15 years is within the domain of our graph, we use interpolation. Find 15 years on the horizontal axis and then draw a vertical line that intersect the line of best fit. Then draw a horizontal line over to the vertical axis.

We can predict that after 15 years of experience, that the income will be approximately $42 000.

When we can't use the graph but require the regression function to make predictions then we have used **Extrapolation.** We have estimated a value **OUTSIDE** the domain of a set of data.

**Example:**

The following data shows the relationship between chirps per second of a striped ground cricket and the corresponding ground temperature.

a) Determine a linear regression model equation to represent this data.

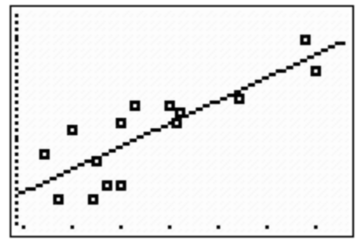
b) Graph the new equation.

c) Extrapolate data: If the ground temperature reached 95º, then at what approximate rate would you expect the crickets to be chirping?

| Chirps/Second | Temperature (º F) |
|---|---|
| 20.0 | 88.6 |
| 16.0 | 71.6 |
| 19.8 | 93.3 |
| 18.4 | 84.3 |
| 17.1 | 80.6 |
| 15.5 | 75.2 |
| 14.7 | 69.7 |
| 15.7 | 71.6 |
| 15.4 | 69.4 |
| 16.3 | 83.3 |
| 15.0 | 79.6 |
| 17.2 | 82.6 |
| 16.0 | 80.6 |
| 17.0 | 83.5 |
| 14.4 | 76.3 |

***Solution:***

a) The linear regression equation is $y = 3.410x + 22.849$

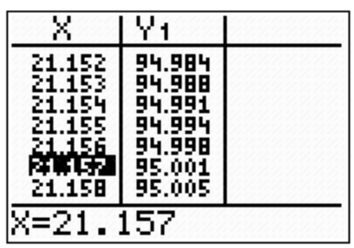


b. Using ZOOM #9 ZoomStat to see the graph

c) *If the ground temperature reached 95º, then at what approximate rate would you expect the crickets to be chirping per 15 seconds?*

**METHOD 1:Using Technology**

Go to TBLSET (above WINDOW) and set the TblStart to 20+ (since the highest temperature in the data set had 20 chirps/second). Set the delta Tbl to a decimal setting of your choice. Go to TABLE (above GRAPH) and arrow up or down to find your desired temperature, 95º, in the Y1 column.

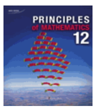

There will be approximately 21.157 chirps/second.

**METHOD 2: Substitution into the linear regression function.**

Since 95 represents the temperature we substitute it into the *y* of the equation.

$$\begin{aligned} y &= 3.410x + 22.849 \\ 95 &= 3.410x + 22.849 \\ 95 - 22.849 &= 3.410x + 22.849 - 22.849 \\ 72.151 &= 3.410x \\ \frac{72.151}{3.410} &= \frac{3.410x}{3.410} \\ x &= 21.16 \end{aligned}$$

There will be approximately 21.16 chips/second.




**Practice Problem:**

Complete “Practising” question 4 on page 302 of your textbook.

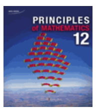
***Solution:***

a. e.g.. The line of best fit has a negative slope. Approximately the same number of points lie above and below the line. There seems to be a strong representation of the data because all the points are very close to the line of best fit.

b) e.g., I estimate that when x = 47, y = 75. I used interpolation, because the point is within the domain of the data.

c. e.g., I estimate that when y = 70, y = 52. I used interpolation, because the point is within the domain of the data.

e.g., I estimate that when x = 15, y = 105. I used extrapolation, because the point is outside the domain of the data.



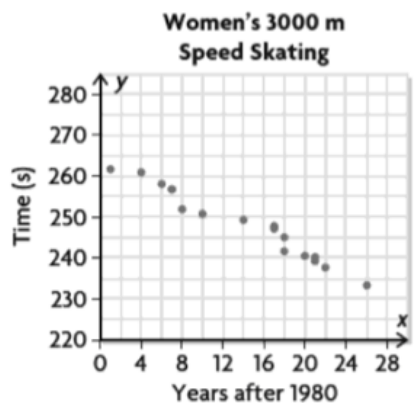

**Practice Problem:**

Complete “Practising” question 5 on page 302 of your textbook.

---

***Solution:***

a) Let x represent the number of years after 1980, and let y represent the record breaking time, in seconds.




b) As the number of years after 1980 increases, the record time in the women's 3000 m speed skating decreases.

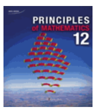
c) The linear regression function is

$$y = -1.147\ldots x + 264.178\ldots.$$

The slope represents the number of second that the record time decreases each year and the y-intercept represents the record time in 1980.

d. e.g., It appears that in 2005, the record breaking time is 235.489 s, or 3:55.49 min

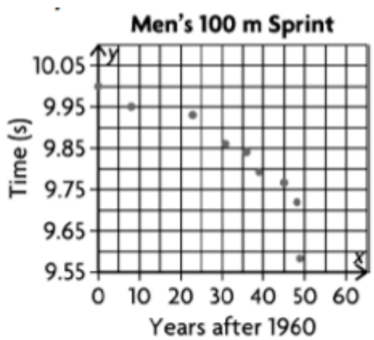
e. Cindy Klassens' actual time in 2005 was 3:55.75 min, which is only about 0.26 seconds higher than my estimate.




**Practice Problem:**

Complete “Practising” question 6 on page 303 of your textbook.

***Solution:***



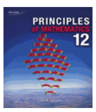

b. Most of the data follows a general pattern, which is that as the years pass, the world-record time will decrease slightly. The main outlier is the record at 49 years, set by Usain Bolt.

c) The linear regression function is $y = -0.006\ldots x + 10.032\ldots.$

The slope represents the amount of time in second that the world-record time will decrease every year and the y-intercept represents the record in 1960.

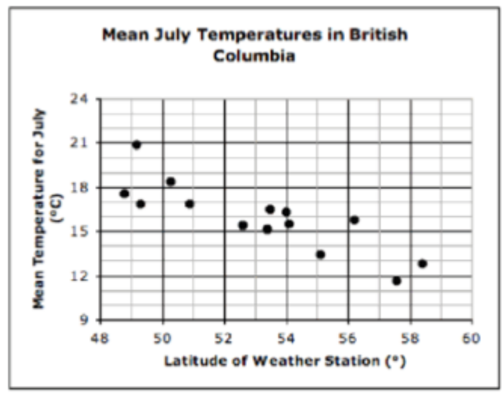
d) e.g., A possible world-record time for 2007 could be 9.72 s.

e) Asafe Powell's time in 2007 was 9.74 s, slightly higher than my estimate.




**Practice Problem:**

Complete "Practising" question 7 on page 303 of your textbook.




***Solution:***

a) e.g., I would expect to see that when the latitude of weather station increases, the mean temperature will decrease.

b) The linear regression equation is

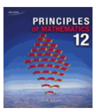
$$y = -0.637\ldots x + 49.730\ldots.$$

c)
$$y = -0.637\ldots x + 49.730\ldots$$
$$y = -0.637\ldots(52.0) + 49.730\ldots$$
$$y = 16.600\ldots$$

At a latitude of 52.0 degrees North, the mean temperature for July will be 16.6 degrees Celsius.

d)
$$y = -0.637\ldots x + 49.730\ldots$$
$$18 = -0.637\ldots x + 49.730\ldots$$
$$-31.730\ldots = -0.637\ldots x$$
$$x = 49.803\ldots$$

A mean July temperature of 18 degrees Celsius can be expected when the latitude of weather station is 49.8 degrees North.



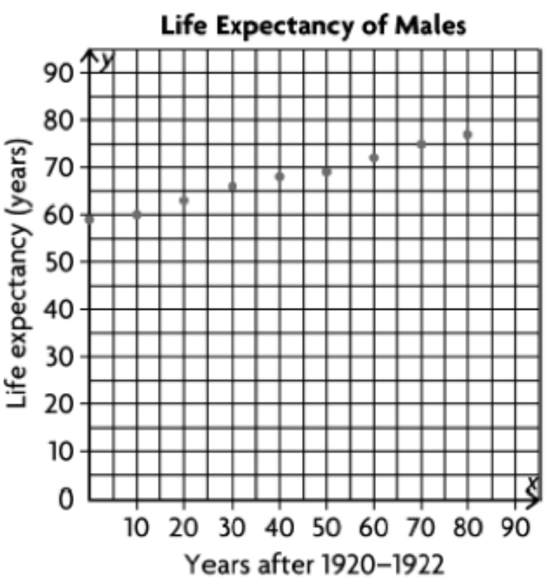
**Practice Problem:**

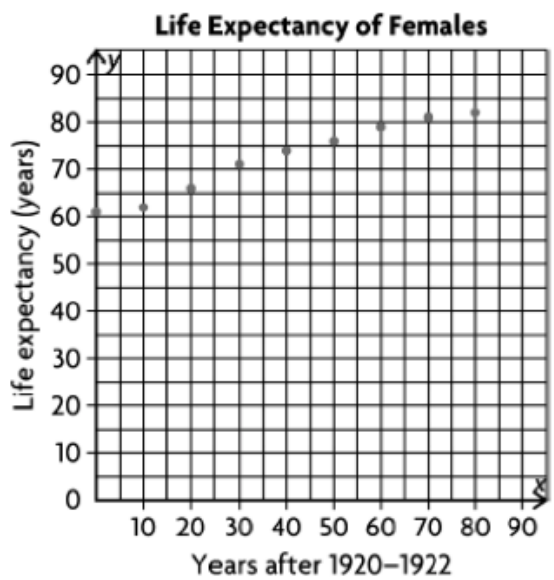
Complete “Practising” question 11 on page 305 of your textbook.

---

***Solution:***

Let x represent the number of years after 1920 - 1922



**b)** The linear regression equation for the male plot is $y = 0.23x + 58.466\ldots$. The linear regression equation for the female plot is $y = 0.286\ldots x + 60.977\ldots$.

**c)** For males:

$y = 0.23x + 58.466\ldots$

$y = 0.23(90) + 58.466\ldots$

$y = 79.166\ldots$

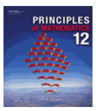
For females:

$y = 0.286\ldots x + 60.977\ldots$

$y = 0.286\ldots(90) + 60.977\ldots$

$y = 86.777\ldots$

In 2010, the life expectancy in Canada will be 79 years for males and 87 years for females.




**Practice Problem:**

Complete “Practising” question 13 on page 305 of your textbook.

***Solution:***

**13. a)** e.g., The easiest method would be to create a scatter plot using the data and then perform a linear regression. Insert the independent variable into the equation for the linear regression and solve for the dependent variable. For example, if a graph has a linear regression with the equation $y = 5x + 1$, then to estimate $y$ when $x = 6$, simply insert $x = 6$ and you get $y = 31$.

**b)** Again, it is best to create a scatter plot using the data and then perform a linear regression. Then insert the dependent variable and rearrange the equation to solve for the independent variable.