**Math 3201**

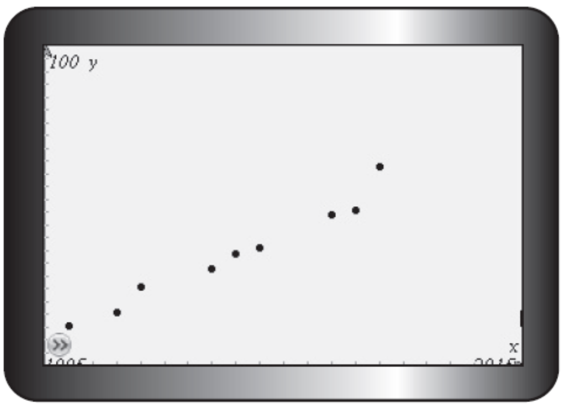
# 5.3 – 5.4 Lines and Curves of Best Fit

## Modelling Data with a Line of Best Fit

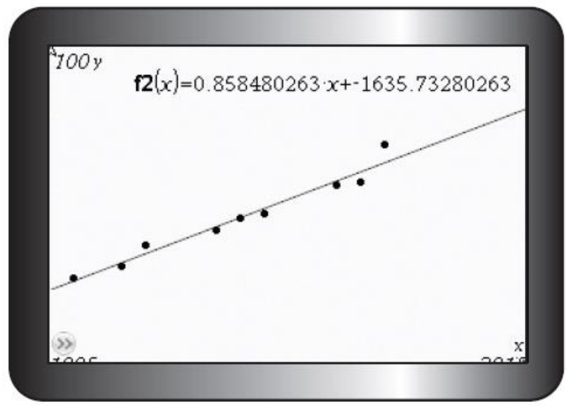
Sometimes when we plot data that represents a real life situation, the points may not perfectly trace out a line, but they may come close to it. In this case, we can draw a **line of best fit** through the data.

**Line of Best Fit**: a line that best approximates the linear trend in a scatter plot.

Scatter Plot




Line of Best Fit




**Regression Function:** a line or curve of best fit, developed through a statistical analysis of data. An equation can be developed for the line or curve.

**Linear Regression:** can be carried out using a graphing calculator to determine the equation of a line of best fit. This equation can then be used to predict other values.

## Estimating Values using Linear Regression

Once an equation is obtained for a line of best fit, we can use that equation to estimate values. There are two estimation processes that we can carry out:

**Interpolation:** the process used to estimate a value within the domain of a set of data, based on a trend. In other words, it involves estimating points that fall in between the data points we already have.

**Extrapolation**: the process used to estimate a value outside the domain of a set of data, based on a trend. That is, we would estimate a point that is beyond or outside the points we already know.

Consider the following table:

| | A | B |
|---|---|---|
| 1 | Years after 1990 | Distance (km) |
| 2 | 6 | 78.04 |
| 3 | 8 | 79.14 |
| 4 | 9 | 81.16 |
| 5 | 12 | 82.60 |
| 6 | 13 | 83.72 |
| 7 | 14 | 84.22 |
| 8 | 17 | 86.77 |
| 9 | 18 | 87.12 |
| 10 | 19 | 90.60 |

If we were asked to estimate the distance at 10 years after 1990, we would be **interpolating** a value, since we would estimate a value that lies in between points that we already know.

If we were asked to estimate the distance at 25 years after 1990, we would be **extrapolating** a value, since we would estimate a value that lies outside the points that we already know.

Whenever a regression model is used to fit a group of data, the range of the data should be carefully observed. Attempting to use a regression equation to predict values outside of this range, extrapolation, is often inappropriate. For example:

- A linear model relates weight gain to the age for young children. Applying such a model to adults, or even teenagers, would not be appropriate, since the relationship between age and weight gain is not consistent for all age groups.

- A cubic model relates the number of wins for a hockey team over time, in years. If students extrapolate too far in the future, their model could yield more wins than is possible in a year.

**Example 1:**
The one-hour record is the farthest distance travelled by bicycle in 1 h. The table below shows the world record distances and the dates they were accomplished.

| 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 |

International Human Powered Vehicle Association

(A) Use technology to create a scatter plot and to determine the equation of the line of best fit.

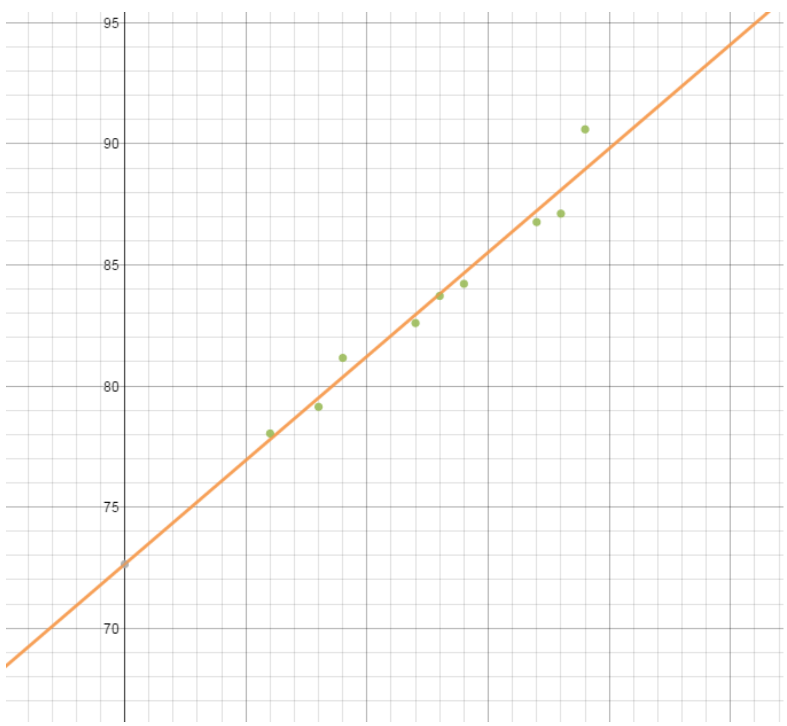
Using Desmos Graphing at [http://www.desmosgraphing.com](http://www.desmosgraphing.com), we will enter the above table and perform a linear regression.

Select the table app, from the drop down menu on the left. Enter the data:

| $x_1$ | $y_1$ |
|---|---|
| 6 | 78.04 |
| 8 | 79.14 |
| 9 | 81.16 |
| 12 | 82.60 |
| 13 | 83.72 |
| 14 | 84.22 |
| 17 | 86.77 |
| 18 | 87.12 |
| 19 | 90.6 |

To perform a **linear** regression in Desmos Graphing, you use the following command:

y_1~mx_1+b, which will appears as: $y_1 \sim mx_1 + b$



Desmos Graphing also calculates the values of the slope, $m$, and the $y$-intercept, $b$:

PARAMETERS

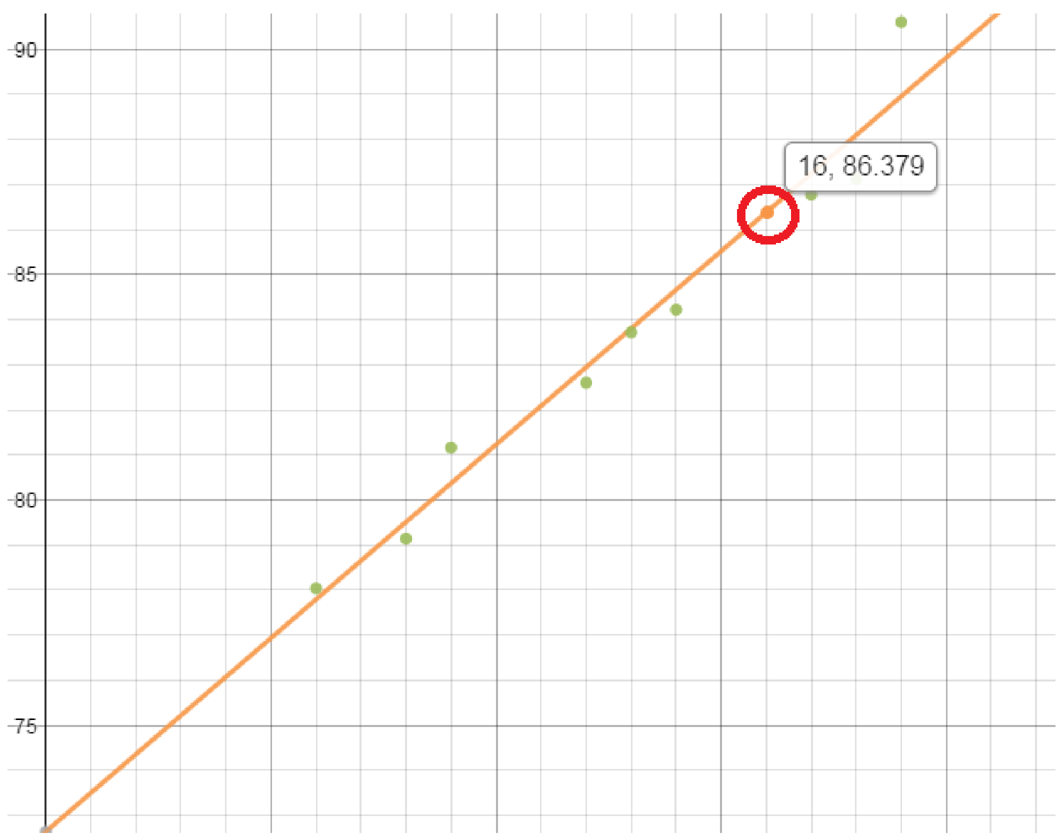
$m = 0.858$ $\qquad$ $b = 72.6$

We can use these values to create the equation of the line that "best fits" these data values:

$y = 0.858x + 72.6$

(B) Interpolate a possible world-record distance for the year 2006, to the nearest hundredth kilometer.

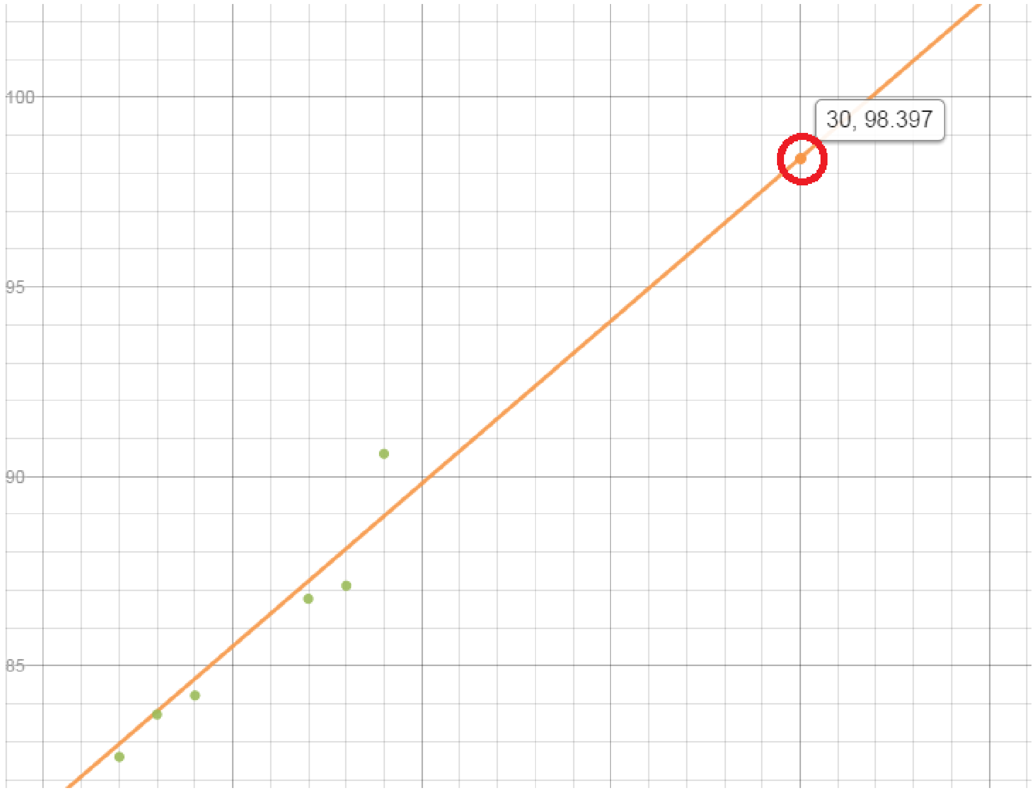
Using the graph generated by Desmos, simple find the $x$-value that corresponds to 2006 and trace the line up to the line of best fit. Click on the line at this point and the app will tell you the $y$-value.




$2006 - 1990 = 16$

86.4 km

(C) Extrapolate a possible world-record distance for the year 2020, to the nearest hundredth kilometer.




98.40km

(D) Use the equation generated by Desmos to calculate a possible world-record distance for the year 2016.

$y = 0.858x + 72.6$

$2016 - 1990 = 26$

$y = 0.858(26) + 72.6 = 94.9$ km

**Curve of Best Fit**

Just as we drew lines of best fit through linear data and performed linear regressions to obtain the equation of the line of best fit, we can do something similar with quadratic and cubic data.

With quadratic and cubic data, we draw a **curve of best fit.**

**Curve of Best Fit:** a curve the best approximates the trend on a scatter plot.

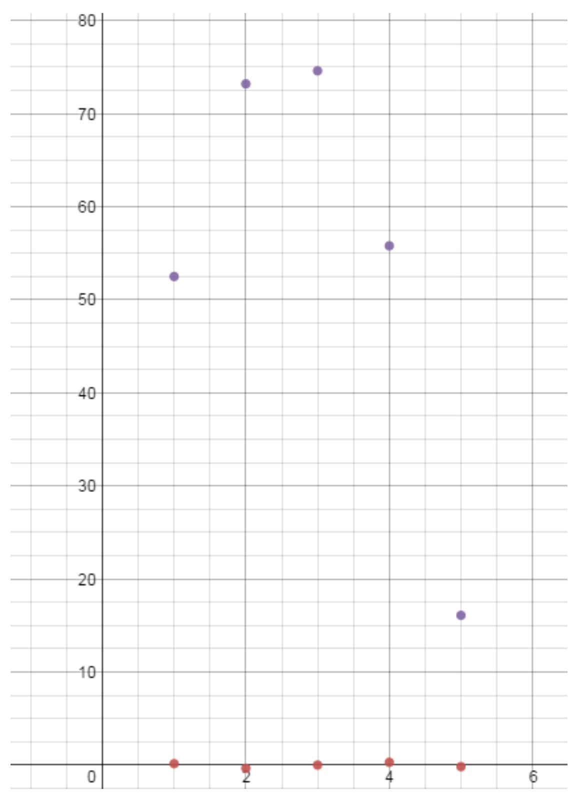
If the data appears to be quadratic, we perform a quadratic regression to get the equation for the curve of best fit. If it appears to be cubic, then we perform a cubic regression.

**Example 2:**
Gene hit a golf ball from the top of a hill. The height of the ball above the green is given in the table below.

| Time (s) | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Height (m) | 52.5 | 73.2 | 74.6 | 55.8 | 16.1 |

(A) Using Desmos Graphing, plot the data points and describe the characteristics of the data.

| $x_1$ | $y_1$ |
|---|---|
| 1 | 52.5 |
| 2 | 73.2 |
| 3 | 74.6 |
| 4 | 55.8 |
| 5 | 16.1 |



(B) Determine the equation of the quadratic regression function that models the data.

The command for a quadratic regression in Desmos is: y_1~ax_1^2+bx_1+c

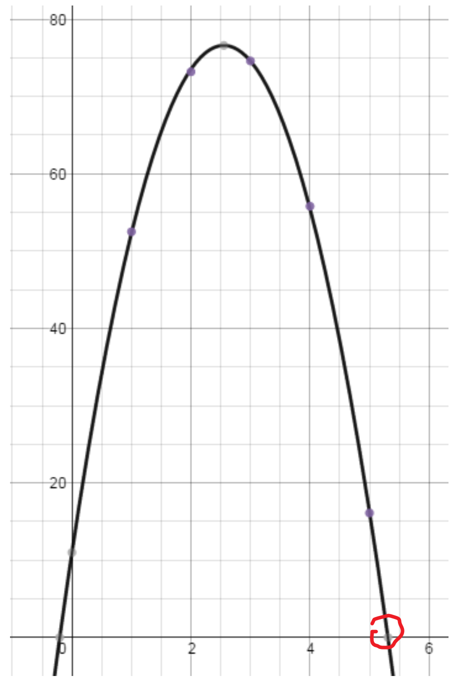
This appears as: $y_1 \sim ax_1^2 + bx_1 + c$

PARAMETERS

$a = -10.1$ $b = 51.4$

$c = 11$

$y = -10.1x^2 + 51.4x + 11$




(C) Use your equation to determine the height of the ball at

i. 0 s

$y = -10.1(0)^2 + 51.4(0) + 11 = 11\,m$

ii. 2.5 s

$y = -10.1(2.5)^2 + 51.4(2.5) + 11 = 76.4\,m$

iii. 4.5 s

$y = -10.1(4.5)^2 + 51.4(4.5) + 11 = 37.8\,m$

(D) When did the ball hit the ground?

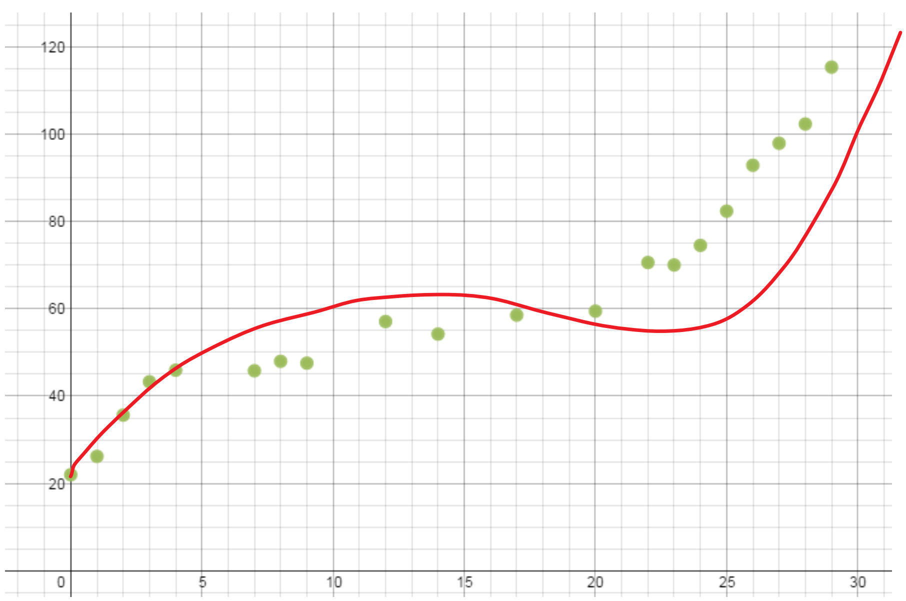
$\sim 5.3\ s$

**Example 3:**
The following table shows the average retail price of gasoline, per litre, for a selection of years in a 30-year period beginning in 1979.

| Years after 1979 | Price of Gas (¢/L) | Years after 1979 | Price of Gas (¢/L) |
|---|---|---|---|
| 0 | 21.98 | 17 | 58.52 |
| 1 | 26.18 | 20 | 59.43 |
| 2 | 35.63 | 22 | 70.56 |
| 3 | 43.26 | 23 | 70.00 |
| 4 | 45.92 | 24 | 74.48 |
| 7 | 45.78 | 25 | 82.32 |
| 8 | 47.95 | 26 | 92.82 |
| 9 | 47.53 | 27 | 97.86 |
| 12 | 57.05 | 28 | 102.27 |
| 14 | 54.18 | 29 | 115.29 |

Statistics Canada

(A) Use technology to graph the data as a scatter plot. What polynomial function could be used to model the data?



(B) Determine the cubic regression equation that models the data. Use your equation to estimate the average price of gas between 1984 and 1985. Would this estimation be interpolation or extrapolation?

PARAMETERS

$a = 0.0123$ $\quad b = -0.465$

$c = 6.3$ $\quad d = 23.5$

$$y_1 \sim a x_1^3 + b x_1^2 + c x_1 + d$$

$y = 0.0123x^3 - 0.465x^2 + 6.3x + 23.5$

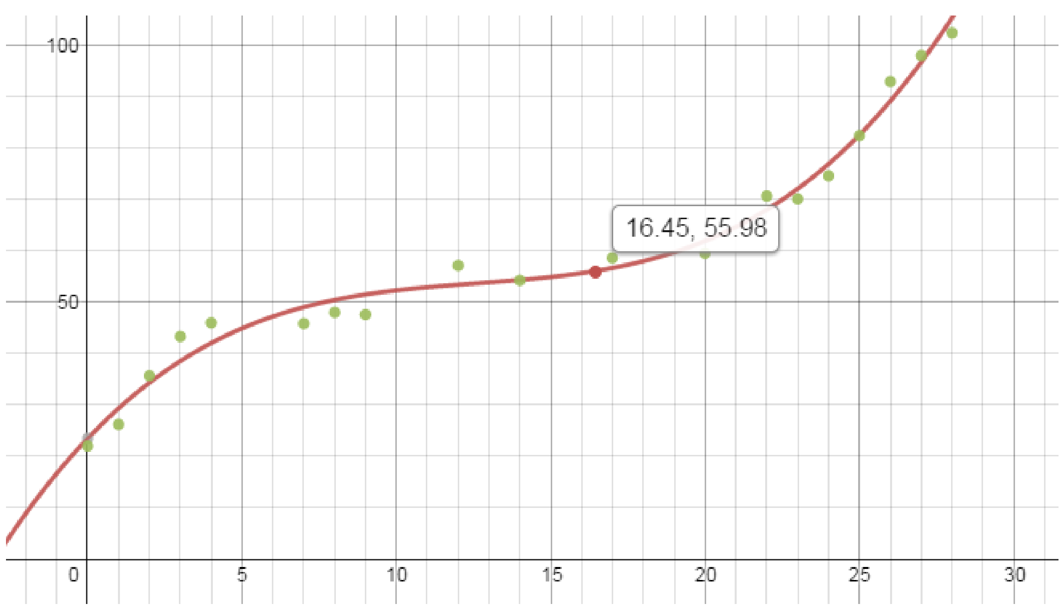
1985 - 1979 = 6 years.

$y = 0.0123(6)^3 - 0.465(6)^2 + 6.3(6) + 23.5$

$y = 47.21$¢/L

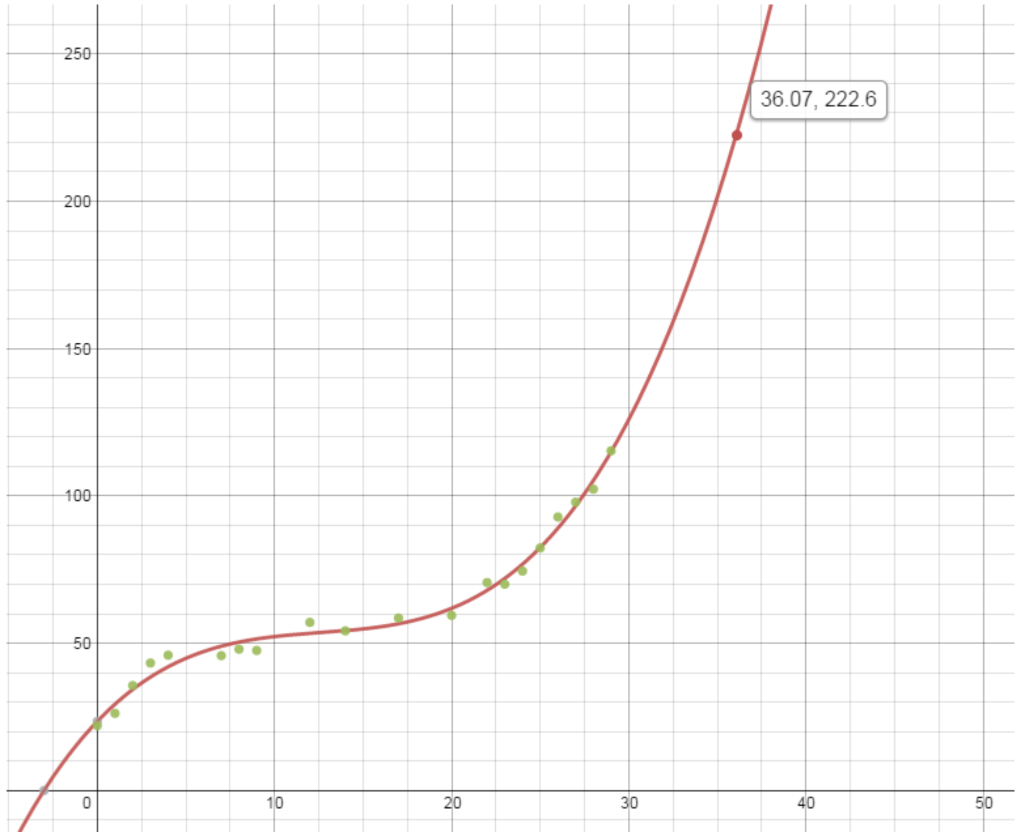
Interpolation

(C) Estimate the year in which the average price of gasoline was 56.0 cents/L.



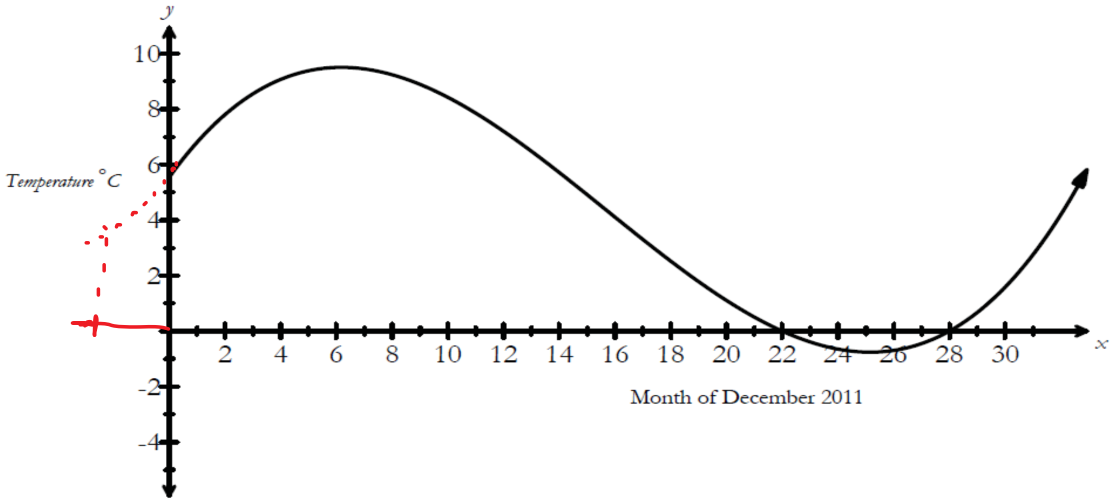

~ 1995

(D) According to this data, what should the price of gas be this year? Is this data a good model for current gas prices?




No. Gas is 94¢/L today.

**Example 4:**
The following graph was used to model the changes in temperature last December. Ask students to answer the following questions:




(A) What would your prediction be for the temperature on Jan $1^{st}$, 2012 ?

4°C

(B) What was the coldest day of the month?

Dec 25th, -1°C.

(C) What might the temperature have been on Nov. $27^{th}$, 2011?

~3 or 4°C

(D) Would you use this model to predict the temperature on Jan. $15^{th}$, 2012? Explain your reasoning.

No. By definition climate changes.

**Textbook Questions:** page 301 - 306 #6, 7, 8, 10
page 313 - 314 #1, 2, 4, 5, 6, 7, 8