NOTES 6.3: Modelling Data with Lines of Best Fit

Lesson Focus: To determine the linear function that best fits a set of data and use the function to solve a problem.

- a ***scatter plot*** is a set of points on a grid used to visualize a relationship or a possible trend in the data
- the ***independent variable*** is the variable or characteristic of the data that is being manipulated
- the ***dependent variable*** is the variable or characteristic of the data that is being observed
  ex. the height of an individual (dependent variable) is correlated to their age (independent variable)
- the independent variable is **always** placed on the horizontal axis of a graph

- if the points on a scatter plot seem to follow a linear trend, then there may be a linear relationship between the independent and dependent variables
- a ***line of best fit*** is a straight line that best approximates the trend in a scatter plot
- we can generate a line of best fit by either:
  1. graphing the data on a grid and then using the edge of a clear ruler to approximate a line that best describes the points
  2. put the data in the List function of your calculator and then perform a regression
- a ***regression function*** is a line or curve of best fit developed through a statistical analysis of the data
- determining the line of best fit with your calculator requires the following procedures:
  1. ***create a scatter plot of the data***
     a) clear all the data from your lists → 2nd/+/4/ENTER
     b) place your data in the list → STAT/ENTER
- if you notice that your lists do not go from L_1 to L_6, use SetUpEditor → STAT/5/ENTER
     c) put the independent variable in L_1 and the dependent variable in L_2
- press ENTER to have the data go into the list
- you must have the same number of elements in each list
     d) your window settings must include both the smallest and largest values for the independent and dependent variables → WINDOW
     e) you can see the scatter plot by turning on your plots →2nd/Y=/ENTER
- make sure Plot1 is **On**, the Type: is a **scatterplot** (1st choice in the 1st row), the Xlist is $\mathbf{L_1}$, the Ylist is $\mathbf{L_2}$ and the Mark: is the first symbol

  2. ***run a regression on the data***
- in order to run a linear regression, use LinReg (ax+b) → STAT/►/4/ENTER

  3. ***graph the regression function with the scatter plot***
- you can paste the regression equation directly into Y_1= by pressing the following keystrokes
  Y=/VARS/5/►/►/ENTER/GRAPH

- the process of ***interpolation*** is used to estimate a value **within** the domain of a set of data, based on a trend
- we can interpolate from our regression equation if we already know the x-value (independent variable)
- press **2nd/TRACE/1** and then enter the known x-value at the **X=** prompt
- the process of ***extrapolation*** is used to estimate a value **outside** the domain of a set of data, based on a trend
- we can extrapolate from our regression equation if we already know the y-value (dependent variable)
- enter our known y-value into Y_2, press **2nd/TRACE/5/ENTER/ENTER/ENTER** and find the intersection

point

Ex #1 The winning times for the men`s 20 km biathlon in the Winter Olympics from 1964 to 2010 (except 2002) are shown in the table below.

Year	1964	1968	1972	1976	1980	1984
Winning Time (min)	80.4	73.8	75.9	74.2	68.3	71.9
Year	1988	1992	1994	1998	2006	2010
Winning Time (min)	56.6	57.6	57.4	56.2	54.3	48.4

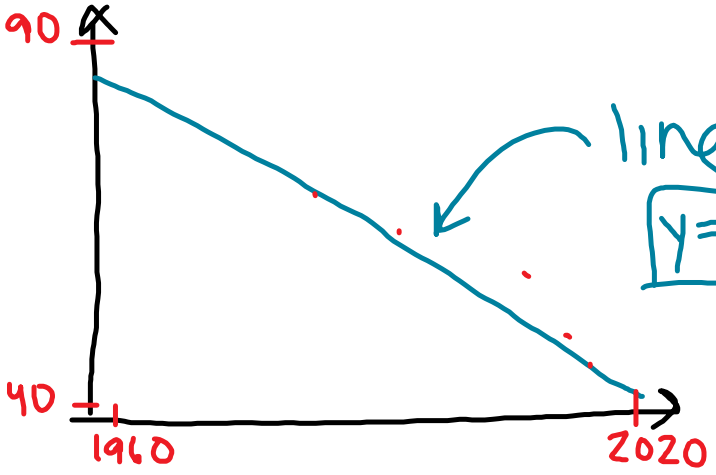
a) Enter the data into your graphing calculator. The independent variable is the year (L_1) and the dependent variable is the winning time (L_2).

b) What Window settings should you use?

X_{min}= 1960, X_{max}= 2020, Y_{min}= 40, Y_{max}= 90

zoom 9

c) Draw a sketch of what you see. Label the axes.



zooms in on the data

line of best fit

Y= Vars 5 ▷▷ ENTER

GRAPH

d) Perform a linear regression of the data. Write the linear regression equation for the scatter plot. Set your calculator to **three** decimals.

$y = -0.682x + 1419.391$

e) Paste the regression equation into Y_1=. Add it to your sketch of the scatter plot in part c).

f) Determine a possible winning time for the event in the 2002 Winter Olympics. Will you use **Value** or **Intersect**? Is this an example of extrapolation or interpolation?

2nd TRACE 1

x = 2002

y = 53 seconds

g) Estimate the possible winning time for the event in the 2014 Winter Olympics? Will you use **Value** or **Intersect**? Is this an example of extrapolation or interpolation?

x = 2014

y = 45 seconds

h) In which year would the winning time have been 65 minutes? Will you use **Value** or **Intersect**?

2nd TRACE 5 → trace along the line to find y = 65

x = 1985

pg 407 # 3-7