Lesson 3.3 – Line of Best Fit

Goal: Analyze two-variable data using a line of best fit

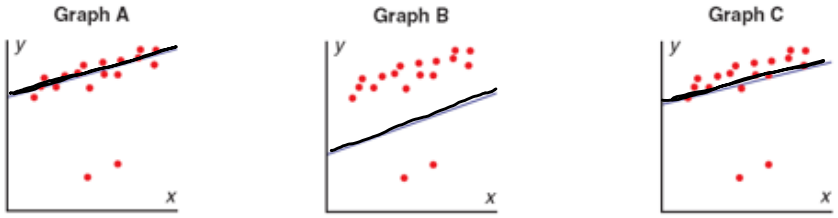
What is a Line of Best Fit (LOBF)?

- A straight line drawn through data points to **BEST** represent a linear relationship between two variables. AKA trend line or regression line
- The weaker the correlation the more difficult it is to draw an LOBF

The Effect of Outliers on the LOBF

- An outlier is a point that lies far away from the main cluster of data
- May be caused by inaccurate measurements or anomalies in the data set (i.e., an unusually tall person)
- The LOBF should reflect all points in the data set, including outliers
- One or more outliers will affect the path of the LOBF

EXAMPLE Analyze each graph below to determine which graph has the line of best fit for the data.



Notice the data in each graph is the same. Only the lines of best fit are different.

Graph A: No good.
This line ignores the two outliers and goes through the middle of the main cluster of data

Graph B: Not good.
This lines gives too much weight to the outliers

Graph C: Good!
This line is slightly lower than the main cluster of data to show that there are two points well below it.

Interpolating and Extrapolating

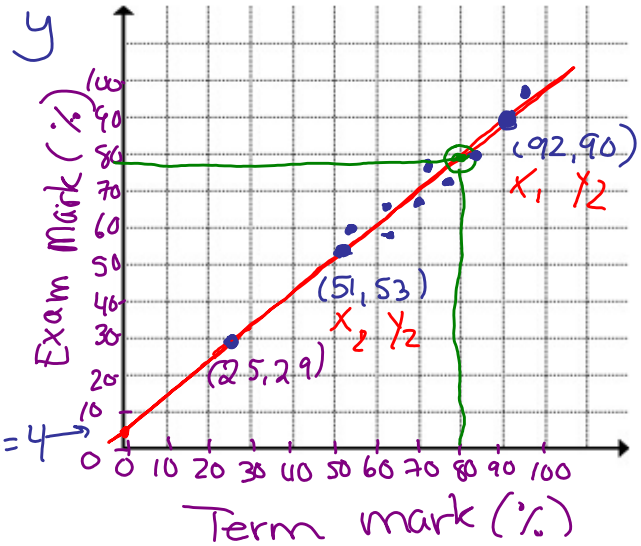
A line of best fit can be used to interpolate or extrapolate values
Estimating values that lie **among** known values on the graph (within the range of data) is interpolating
Predicting values that lie **beyond** known values is extrapolate

Using a Line of Best Fit to Make Predictions

EXAMPLE Below are pre-exam term marks and exam marks for some students in a Grade 12 math class.

Term Mark (%)	84	76	70	95	92	61	25	55	51	73	62
Exam Mark (%)	80	72	68	96	90	58	29	60	53	77	67

a) Graph the data and draw the line of best fit.



b) Determine the equation of the line of best fit.

strong positive correlation

$y = mx + b$ (m: slope, b: y-intercept)

$b = 4$

$m = \frac{\text{rise}}{\text{run}} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{90 - 53}{92 - 51} = \frac{37}{41} = 0.9$

$y = 0.9x + 4$ (x)

$m = \frac{53 - 29}{51 - 25} = \frac{24}{26} = 0.92$

c) Use the **graph** to estimate the exam mark of a student with a pre-exam term mark of 80%

Approximately 78%

(Interpolation)

$y = 0.9(80) + 4$

$= 76\%$

d) Use the **equation** to predict the exam mark of a student with a term mark of 10%

$y = 0.9(10) + 4$

$= 13\%$

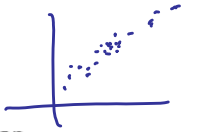
(Extrapolation)

Deciding if a Correlation is Linear

Not all relationships between two variables are linear. We will learn more about non-linear models later.

A linear model may be unreliable if:

- There are too few pieces of data (small sample size)
- The range of data is too narrow
- There does not appear to be any correlation or the data do not appear to be linear
- There are outliers



Practice: Page 153 #1, 3, 5 – 8, 10