MAP4C1

Unit 3: Two-Variable Statistics

Learning Goals: We are learning to...

- 1. Compare one and two variable data and explain the different between them
- 2. Identify if a given situation involves one or two variable data and determine the best graph to use when representing this data.
- 3. Create, use and interpret scatter plots involving two-variable data
- 4. Analyze scatter plots to determine correlation and make connections
- 5. Analyze two-variable relationships to determine if there is a reasonable cause between two variables
- 6. Analyze two-variable data using a line of best fit
- 7. Determine the line of best fit for a given set of data and use interpolation and extrapolation to analyze the data
- 8. Organize, interpret and analyze secondary source data using spreadsheets

Day	Торіс	Expectations, Learning Goals	Practice/Homework	
1	3.1 One and Two Variable Data	LG 1, 2 D1.1	p.133 #1-5, 8	
2	3.2 Scatterplots	LG 3, 4, 5 D1.4, 1.7, 1.8	p.142 p.1-6, 8, 10, 12	
3	3.3 Line of Best Fit	LG 6, 7 D1.5, 1.6, 1.7	p.153 #1, 3, 5-8, 10	
4	3.4 Analyzing Data Using a Spreadsheet	LG 8 D1.3	p.158 #1-5	
5	Review		See review package	
6	Unit 3 Quest			
	Date:			

Subject to change based on school activities and class needs

Learning Goals: I am learning to...

- Compare one and two variable data and explain the different between them
- Identify if a given situation involves one or two variable data and determine the best graph to use when representing this data.



Recall Key Terminology:

Variable	
	•
	•
One-Variable Data	•
	•
	•
	•
Two-Variable Data	•
	•
Correlation	

In order to identify situations of one- or two-variable data, ask yourself the following questions:

- 1. What is the information measuring?
- 2. How can the information be displayed?
- 3. How can the information be analyzed?

Example 1: State whether each situation involves one-variable or two-variable data. **Justify** your answer.

a) Noah researches the annual hours of sunshine in Canadian cities.

b) A study compares the amount of time spend watching TV with the amount of time spent reading.

Types of Data

There are 3 main types of data that can be collected:

Categorical Data	Discrete Data	Continuous Data		

Types of Graphs

There are 4 main ways in which data can be displayed in a graph

Bar Graph	Histogram	Pie Chart	Scatter Plot

In order to identify which graph to use, ask yourself the following questions:

- 1. Is it one- or two-variable data?
- 2. What type of data is it?

Example 2: For a class project, Danny surveyed students about their part-time jobs. The data is shown in the table below.

a) What type of graph would be best to show how many hours a student worked on the weekend?

Student	Weekend hours	Weekday Hours
Anya	5.0	12.5
Ellen	8.0	12.0
Fiona	17.0	8.0
Aaron	0.0	16.5
Leila	10.0	16.0
Mason	9.5	8.0
Petra	15.0	6.0

b) What type of graph would be best to show a possible relationship between the number of hours worked on the weekend vs. weekday?



1. a) Does each graph illustrate one-variable or two-variable data?

2. a) Does each table illustrate one-variable or two-variable data?

i)	Household	Number of TVs	ii)	Student	Dollars
	size	in household		Anne	15
	1	2		Lars	25
	2	3		Mason	5
	3	3		Thom	20
	4	3		Riaz	25
	5	4		Loni	10

III)	Candle burn time (h)	10	15	25	40
	Cost (\$)	7.49	10.99	15.49	19.99

- **b)** Choose one table from part a. How did you decide whether the table showed one- or two-variable data?
- **3.** Identify the two variables in each situation.
 - a) The more purchases made with a credit card, the more reward points earned.
 - **b)** Anthropology students read an article that claimed that people with greater brain mass have higher IQs.
 - c) Across Ontario, the mosquito population remained low due to the lower than average rainfall.

- 4. a) State whether each situation involves one-variable or two-variable data.
 - i) Marcus calculates the median mark of the class on an exam.
 - ii) An article discusses the possible link between prolonged cell phone use and the increased probability of brain cancer.
 - iii) A classroom survey shows that 70% of the students plan to attend university, 15% plan to attend college, 10% are going directly into the workplace, and 5% are uncertain.
 - b) Choose one situation from part a. Explain how you decided whether the situation involved one- or two-variable data.

The median is the middle value when data are arranged in numerical order. If there are two middle values, their mean is the median.

5. What type of graph would you use to display the data in each table? Justify your choices.

a)	Number of Sit-Ups Students Can Do in 1 min								
	Number of sit-ups	0–9	10–19	20-29	30–39	40-49			
	Frequency	0	3	5	8	4			

b)

Land Area of Selected Provinces/Territories								
Province or territory	Alberta	Manitoba	Ontario	Quebec	Nova Scotia	PEI	Nunavut	
Land area (1000 km ²)	642	554	918	1365	53	6	1936	

c)

	Ages of Selected Students by Grade								
Grade	9	11	12	9	10	10	11		
Age	15	18	18	14	15	16	16		

8. A teacher surveyed her students about how long they had studied for a test. a) Which graph displays two-variable data? Justify your answer.







- b) Which graph provides information about a possible relationship? What variables does the relationship involve?
- c) For each graph, write a question someone could answer using the data in the graph. Answer your questions.

Answers:

1.	a)	i) One-variable data
		ii) One-variable data
		iii) One-variable data
		iv) Two-variable data
2.	a)	i) One-variable data
		ii) One-variable data
		iii) Two-variable data
3.	a)	Purchases made and reward points earned
	b)	Brain mass and IQ
	c)	Mosquito population and average rainfall
4.	a)	i) One-variable data
		ii) Two-variable data
		iii) One-variable data
5.	a)	Histogram
	b)	Circle graph
	c)	Scatter plot
8.	a)	Study Time and Test Scores
	b)	Study Time and Test Scores provides information
	-,	about a possible relationship. The variables
		involved are Study time and Mark on test.

Learning Goals: I am learning to...

- Create, use and interpret scatter plots involving two-variable data
- □ Analyze scatter plots to determine correlation and make connections
- □ Analyze two-variable relationships to determine if there is a reasonable



cause between two variables

In lesson 3.1, we learned that a scatter plot is used for two-variable data to show how two pieces of information can be collected and compared to see if there is a relationship between the two variables.

Part A: Creating a Scatter Plot

Example 1: Shayna recorded the cost of customers restaurant bills and the tip amount left. The data she collected is displayed in the table below. Create a scatter plot to display the data collected.

\$ Bill	\$25.15	\$38.49	\$19.27	\$49.66	\$32.45	\$72.14
\$ Tip	\$4.00	\$5.50	\$2.50	\$9.00	\$5.00	\$14.00

Your scatter plot should include the following:

- ✓ Title
- ✓ Axes labels

Independent variable (_____) relies on nothing
 Dependent variable (_____) relies on the independent variable

- ✓ Appropriate scale
- ✓ Plot the (x, y) coordinates
- ✓ Draw a line of best fit (LOBF)
- \checkmark State correlation:

Part B: Interpreting a Scatter Plot

Example 2: Jay researched estimates for a job painting his house. The scatter plot below shows the results he collected.

- 1. Which two companies will take the longest to complete the job? Which of these is the cheaper option?
- 2. Which two companies charge the same amount?



3. Why might you pick company E?

4. Why might you pick company B?

Hiring a Painting Company



Part C: Analyzing a Scatter Plot

Correlation measures the strength of a relationship and the direction in which the relationship occurs.



Example 3: Classify each of the following scatter plots based on their correlation. Note: Your classification should include both direction and strength



Example 4: Davis conducted an experiment comparing a person's leg length and how long it takes to walk 100m. The data he collected is shown in the scatter plot below.

a) What sort of relationship does the graph suggest between leg length and time take to walk 100m?



- b) Use the scatter plot to estimate the time it would take a person with a leg length of 85 cm to walk 100m?
- Leg Length and Time Taken to Walk 100 m



c) How might Davis make the results of his experiment more reliable?

Part D: Cause and Effect Relationships

Observing the relationship between two variables <u>does not always mean that one variable</u> <u>causes a change in the other variable</u>. Other factors could be involved to cause the relationship, or the correlation could be a coincidence. Some relationships are obvious, but others may have a common cause to both variables.

Example 5: State whether the claim made in each situation is reasonable. If not, determine if there is a common cause, or if the relationship is coincidental.

- a) A scientific study showed a negative correlation between aerobic exercise and blood pressure. It claimed that the increase in aerobic activity was the cause of the decrease in blood pressure.
- b) Mila discovered a positive correlation between ice cream sales and the number of drowning incidents. She then warned all her friends not to eat ice cream if they intended on going swimming.

c) Since the 1950's the concentration of carbon dioxide (CO₂) in the atmosphere has be increasing. Crime rates in most countries has also increased during this time period. A newspaper reports that the increase in CO₂ levels in the atmosphere causes people to commit more crimes.

<u>3.2 Scatter Plot Homework</u>

- 1. a) What does each point show?
 - b) Which child wears the smallest shoe? The biggest shoe?
 - c) Which two children are the same age?
 - d) Which two children wear the same shoe size?
- Identify the dependent variable in the scatter plot
 Age in question 1. If you do not think there is a dependent variable, tell why.
- **3.** a) Which company has the lowest moving cost?
 - b) Which two companies use the largest trucks?
 - c) Which two companies use the smallest trucks?
 - d) Based on the scatter plot, how many different truck sizes are there?
 - e) Which company would you use for a small load? For a very large load?







- **4.** Identify the dependent variable in the scatter plot in question 3. If you do not think there is a dependent variable, explain why.
- 5. Which building is represented by each point in the scatter plot?



- **6.** For each situation, state whether you think the two variables would have a positive correlation, a negative correlation, or no correlation.
 - a) Cost of a restaurant bill and the amount left as a tip
 - b) Blood pressure reading and IQ
 - c) Number of applicants for a job and probability that you will get the job
 - d) Speed of current and time taken to travel upstream
 - e) Number of kilometres driven and price of gas per litre

- **8.** For each part, state whether you think the two variables will have any
 - correlation. If you think a correlation exists, describe it briefly.
 - a) Summer temperatures and sales of bottled water
 - b) Price of gasoline and number of people who go to movies
 - c) Price of gasoline and number of day trips people take
 - d) Cost of tuition and number of students who apply to college
- **10.** Each graph shows the time taken for two race cars to travel around an oval track for a few practice laps.



- a) How do these graphs show that there are two cars?
- b) How many laps are shown in each graph? What assumption are you making?
- c) Which graph shows both cars maintaining a fairly constant lap time?
- d) In each graph, which car has the greater average speed?
- e) Suppose it starts to rain, making the track slippery and forcing the cars to slow down. Which graph would best show this? Explain.
- **12.** This table shows the hourly cost of heating a pool with a given surface area.
 - a) Using the table, describe what happens to the hourly cost as the surface area increases. What does this suggest about the trends you might see in a scatter plot of these data?
 - b) Create a scatter plot for these data. How does it compare to your prediction in part a?

100	
200	50
400	102
500	127
800	206
1200	310
1500	390

Hourly cost

(cents)

24

Surface area

(sq. ft.)

100

12. a) The hourly cost increases as the surface area increases. The points on the scatter plot should go up to the right. There should be a positive linear correlation



- Yes, the scatter plot shows a positive correlation C) between the two variables.
- 4944 ¢ or \$49.44

- c) Does the scatter plot reveal a correlation between the two variables? If so, describe it.
- d) About how much would it cost to heat a 20 feet \times 40 feet pool for 24 h?



Û

Ð

N e i

Company A for a small load; company E for a anies A and B use the smallest truck Companies E and F use the largest trucks different truck sizes load Company A large erv (e Â ŝ f e

Point	Building Number
A	s

9

В 0

Moving cost is the dependent variable;

4

۰.

Positive correlation ositive correlation correlation

œ.

(e) a) Â ົບ

- Negative correlation No correlation
- There are two sets of points plotted Negative correlation Ð a)

₫.

- Car A and car B are also labelled on the legend. â
- The two cars use the same track each time.
 - Graph 3
- greater average speed. Car A has the () () ()

Graph 1

2

4

3

D

Answers

Shoe size of a child relative to her/his age

a) Â

÷

Learning Goals: I am learning to...

- □ Analyze two-variable data using a line of best fit
- Determine the line of best fit for a given set of data and use interpolation and extrapolation to analyze the data



What is a line of best fit?

A line of best fit (LOBF) is used when comparing two-variables in a scatter plot. It best

represents a ______ relationship and is a ______ line.

A line of best fit will vary depending on the scatter plot. In general, it can be said that the weaker the correlation, the harder it is to make a line of best fit, since the relationship is loosely linear.

Outliers

An outlier on a LOBF represents a point that is ______ from the other data points.

Outliers can occur for a number of reasons, including:

- Inaccurate measurements
- Anomalies in the data (i.e. You are surveying people about their height and you survey the world's tallest man)

The LOBF you make should reflect all points in your data set, including any outliers. The more outliers you have in a data set though, the greater the impact on your LOBF.

Example 1: Given the three scatter plots below, determine which has the best LOBF and explain why.



Graph A:

Graph B:

Graph C:

Interpolating and Extrapolating

Once a line of best fit has been made, it can be used to interpolate or extrapolate values.

- Estimating values within the set of data is known as ______
- Predicting values <u>outside</u> the set of data is known as ______

Example 2: The table below summarises a Grade 12 math class term marks with the final exam marks.

Term Mark (%)	84	76	70	95	92	61	25	55	51	73	71
Exam Mark (%)	80	72	68	96	90	58	29	60	53	77	73

- a) Graph the data and draw a line of best fit.
- b) Determine an equation for the line of best

fit.



c) Use the graph to approximate the exam mark for a student with a term mark of 80%

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

Is a correlation linear?

Sometimes the relationship between two variables may not appear to be linear. We will look at these models later in the course.

Linear models are not always the most reliable in the following situations:

- There is not enough data (too small sample size)
- The data is clustered together
- There does not appear to be any correlation
- There are multiple outliers
- The general shape of the data does not appear to be linear.

If a relationship between two variables does not appear to be linear, DO NOT force it to be. Simply state that the overall relationship is non-linear.

Practice! Graph the given set of data.

Height (cm)	Shoe Size
182	9
178	9
167	8.5
168	7
175	10
178	10
168	8
178	11.5
172	8.5
157	6.5
160	7.5
170	9.5

\square										
\square										
\square										
\square										

- 1. Describe the correlation between the variables.
- 2. Draw a line of best fit.
- 3. What would you expect the shoe size of a person who is 173 cm tall?

4. How could you make the results more reliable?

3.3 Line of Best Fit Homework

1. For each scatter plot, select the line of best fit. Justify each choice.



- **3.** Use this graph to identify whether you would use interpolation or extrapolation to predict each value.
 - a) Volume of water in a hot tub that requires 10 g chlorine.
 - b) Mass of chlorine needed for a 500 L hot tub.
 - c) Mass of chlorine needed for a 35 000 L hot tub.
 - d) Volume of water in a hot tub that requires 18 g chlorine.





- 5. The line drawn in this graph passes through the points so that half the points lie above the line and half the points lie below the line.
 - a) Why is this line not the line of best fit?
 - b) Describe how the line of best fit would look for these data.



6. Describe the problems with drawing the line of best fit for these data.



- **7.** a) For each scatter plot, describe the relationship between *x* and *y*.
 - **b)** Would you model each relationship with a linear or non-linear model? Justify your answers.





- a) Use the line of best fit for this scatter plot to make each prediction.
 - i) Number of games played by a player who averages 10 shifts per game
 - ii) Number of games played by a player who averages 5 shifts per game
 - b) Which of your predictions in part a do you think is more reliable?





10. Assessment Focus Use the table of data about life expectancy of Canadian males.

Life Expectancy at Birth of a Canadian Male										
Birth year	1920	1930	1940	1950	1960	1970	1980	1990		
Life expectancy at birth (years)	59	60	63	66	68	69	72	75		

- a) Create a graph and draw the line of best fit for these data.
- b) Find an equation of the line of best fit.
- c) Estimate the life expectancy of a male born in 1975.
- d) Predict the life expectancy of a male born in 2000.
- e) Predict the birth year of males with a life expectancy of 80 years.
- f) Write a question someone could answer using your graph. Prepare an answer for the question.



- a) The line of best fit is not in agreement with the trend of the points on the scatter plot.
 - b) The line of best fit would have a negative slope and lie close to the four points.
- There is no clear pattern to these data. If a weak correlation exists, it is difficult to determine whether it is positive or negative.
- 7. a) i) Fairly strong positive correlationii) Strong positive correlation
 - iii) Fairly strong negative correlation
 - Weak positive correlation followed by a fairly strong positive correlation, then a strong positive correlation
 - b) i) Positive linear
 - ii) Non-linear
 - iii) Negative linear
 - iv) Non-linear
- 8. a) i) About 29 games ii) About 11 games

3.4 Analyzing Using a Spreadsheet

For this activity, we will be using Microsoft Excel or Google Sheets. You will also need to access the class website (link was sent to your gapps account). Make sure to **follow all instructions carefully. ACTIVITY I**

- □ Go to Google classroom
- □ Locate and download the file: Lesson 3.4.xls
- □ Open the file. You may wish to save this on your drive.
- □ Locate the sheet **Daylight Hours.** This spreadsheet shows the latitudes of different locations and the number of daylight hours on August 15

Question 1: From the data, describe the relationship between latitude and hours of daylight.

- □ Highlight cells A3 to B12
- □ Insert a Scatter Plot from the Insert Ribbon choose the first graph with no lines
- Right-click on the Series1 legend and select Delete
- Select the Chart title and Axes titles icons from the Layout Ribbon to enter titles and labels as shown
- Right-click on the horizontal axis and select Format Axis
- □ Change the *Minimum* and *Maximum* options to a *Fixed* setting of 0 and 90, respectively



Daylight Hours on August 15

- Right-click on the vertical axis and select Format Axis
- □ Change the *Minimum* and *Maximum* options to a *Fixed* setting of 0 and 24, respectively and change the *Major Unit* to a *Fixed* setting of 4

Question 2: From the scatter plot, describe the correlation between the variables. Compare this with your description from question 1.

Question 3: Do you think a linear model would represent this data well? Explain your thinking.

- Right-click on the (+) icon on the right of the graph and select Add Trendline
- □ Select the Linear Trend/Regression Type
- Select the Display Equation and Display R-Squared value checkboxes

* The R² value ranges from 0-1 and tells us how well the data fits the model/trendline. In general, the higher the R² value, the better the fit, however, this is not true for all cases.

Question 4: What is the equation of the line of best fit?

Question 5: Do you think the line does a good job of representing this data? Would the linear model provide reliable estimates of daylight hours? Justify your answer.

- □ Go to B13.Type in: =average(B3:B12)
- \Box Go to B14. Type in: =median(B3:B12)

Question 6: Record the mean (average) and median of the data set.

Format Trendline	8							
۵ 🔶 💼								
 Trendline Options 								
Exponential								
Linear								
Logarithmic								
📈 🔵 Polynomial	Order							
🤳 🔵 Power								
Moving Average	Per							
Trendline Name								
Automatic	Linear (Series1)							
Oustom								
Forecast								
Forward	0.0 periods							
Backward	0.0 periods							
Bet Intercept								
Display Equation on chart								
Display R-squared value on chart								

Question 1: From the data, describe the relationship between latitude and hours of daylight.

Question 2: From the scatter plot, describe the correlation between the variables. Compare it with your description of the relationship from question 1.

Question 3: Do you think a linear model would represent the data well? Explain your thinking.

Question 4: What is the equation of the line of best fit?

Question 5: Do you think the line does a good job of representing these data? Would the linear model provide reliable estimates of daylight hours? Justify your answer.

Question 6: Record the mean (average) and median of the data set.

ACTIVITY II

- □ Go to Google classroom
- □ Locate and download the file: Lesson3.4.xls
- □ Open the file. You may wish to save this on your drive.
- □ Locate the sheet **Snow Rain.** The spreadsheet shows the "weather normals" from Environment Canada, that represent typical weather data for different locations. The current "normals" are based on data collected from 1971 to 2000.

Question 1: Do the two variables appear to be related? If so, describe the relationship. If not, explain why not.

Create a scatter plot for the data. In Excel, the independent variable always defaults to the first column

Question 2: Describe any correlation you see. Does the graph support your answer to question 1?

Add a line of best fit to the graph

Question 3: How well do you think the line of best fit represents the data? Justify your answer.

Question 4: Petawawa receives an average of 228.5 cm of snow each year. Based on the line of best fit, what would you expect the average annual rainfall to be in Petawawa? How close was the prediction to the actual average of 615.9 mm?

ACTIVITY II: Response

Question 1: Do the two variables appear to be related? If so, describe the relationship. If not, explain why not.

Question 2: Describe any correlation you see. Does the graph support your answer to question 1?

Question 3: How well do you think the line of best fit represents the data? Justify your answer.

Question 4: Petawawa receives an average of 228.5 cm of snow each year. Based on the line of best fit, what would you expect the average annual rainfall to be in Petawawa? How close was the prediction to the actual average of 615.9 mm?

ACTIVITY III

- □ Go to Google classroom
- □ Locate and download the file: Lesson 3.4.xls
- □ Open the file. You may wish to save this on your drive.
- Locate the sheet Jan Temp. Latitude and longitude describe the location of places on earth. Latitude describes the location in degrees north or south of the Equator. Longitude describes the location in degrees east or west of the Prime Meridian.

Question 1: Looking at the data, predict/explain whether there is a relationship between the mean January temperature of a North American city and the city's Latitude position north of the Equator.

□ Create a scatter plot for January temperature and latitude.

Question 2: Describe the correlation.

□ Insert a trendline for the scatter plot and determine its equation and R-squared value.



Question 3: Looking at the data, predict/explain whether there is a relationship between the mean January temperature of a North American city and the city's Longitudinal position.

□ Create a scatter plot for January temperature and longitude. To do this, you will have to highlight values from B2 to B51 and then, while holding the Ctrl key, highlight the values from D2 to D51

Question 4: Describe the correlation.

□ Insert a trendline for the scatter plot and determine its equation and R-squared value.

Question 5: Draw a conclusion about the effect of longitude and latitude on temperature.

ACTIVITY III: Response

Question 1: Looking at the data, predict/explain whether there is a relationship between the mean January temperature of a North American city and the city's Latitude position north of the Equator.

Question 2: Describe the correlation.

Question 3: Looking at the data, predict/explain whether there is a relationship between the mean January temperature of a North American city and the city's Longitudinal position.

Question 4: Describe the correlation.

Question 5: Draw a conclusion about the effect of longitude and latitude on temperature.

Mid-Chapter Review

- **1.** Does the situation illustrate one-variable or two-variable data? Explain.
 - a) Students Competing in a Math Contest



- b) Students in a science class measured the height of their plants each week for 12 weeks.
- **3.2 2.** State whether you think the variables in each situation would have a negative correlation, a positive correlation, or no correlation.
 - a) Driving speed and time to travel 100 km
 - b) Size of a house and its interior temperature
 - c) One's age and the number of colds one's had
 - d) Cost of gasoline and fuel efficiency of a vehicle
- **3.3** Which linear model better represents the data? Explain your choice.



4. This table compares the parking facilities of several large companies.

3.2

3.3

Available Land and Parking Capacity for Various Companies							
Acres of land	Parking spaces						
2.0	145						
1.5	160						
4.0	500						
1.0	95						
5.0	600						
4.0	425						
2.0	550						
3.0	280						

- a) Create a scatter plot of the data.
- **b)** Describe any trends you see.
- **5.** Use the table of data on tire pressure.
 - a) Graph the data; draw a line of best fit.
 - **b)** Describe the correlation.
 - c) Predict the pressure at:
 - i) 70°F ii) 40°F

Tire Pressure and Temperature								
Outside temperature (°F)	Tire pressure (psi)							
58	35							
79	38							
63	36							
61	36							
85	39							
55	34							
74	37							
88	40							