

U3D2 MAP 4CI

Scatter Plots

Date: _____

Scatter Plots represent _____.

Scatter plots may indicate a _____ between the two _____.

In two variable data situations, one variable may be _____ on another: in other words, its value _____ according to the _____ variable.

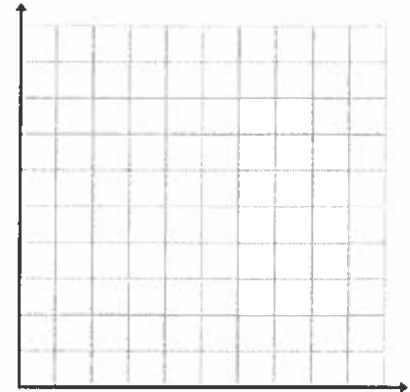
For example, the value of a car depends on _____.

So the independent value is _____

And the dependent value is _____

To plot a scatter plot the dependent variable goes on the _____.

And the independent variable goes on the _____.



CORRELATION DOES NOT MEAN CAUSATION

Example : Daylight hours and Temperature - In winter daylight hours are shortened and temperature is lower (positive correlation). But the low temperature is not caused by the low number of daylight hours. It is caused by the angle the earth makes relative to the sun.

Correlation: _____

Positive Correlation: _____

Negative Correlation: _____

What type of correlation would you expect for the value of a car over time?

Considering Possible Cause and effect. Observing a relationship does not mean that the one variable causes a change in the other. Other factors can be involved or the correlation could be a coincidence.

- 1) State whether the claim in each situation is reasonable.
 - 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) A positive correlation was discovered between the gas price and the average monthly temperature. She concluded that the temperature determines the price of gas.
- 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) A Person's age and the number of colds they have had.
 - d) Price of gasoline at the pump and the fuel efficiency of a vehicle.

U3D2

Do the following variables show positive, negative or no correlation.

- Babies crying and being held by their parents
- Average Daily Temperature and Distance from the North Pole
- Average Daily Temperature and Distance from the Equator
- Speed and the time it takes to drive a given distance.
- Distance driven and the amount of fuel consumed.

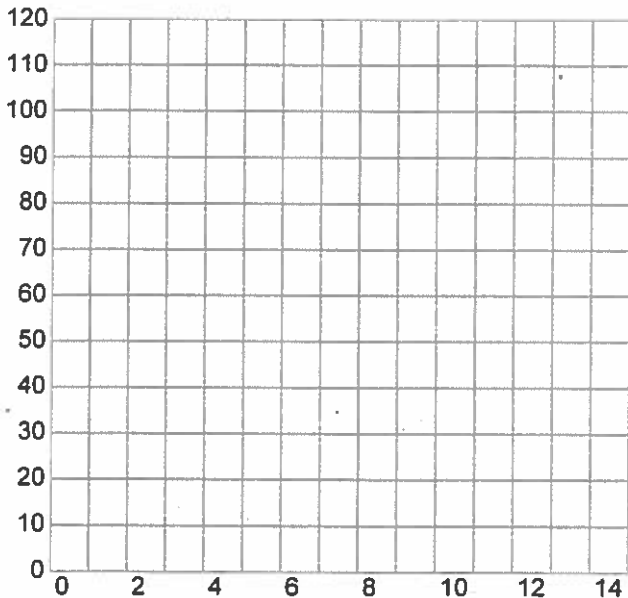
Example: Canadians consume a large amount of sodium. Although some sodium is needed to control blood volume and to help cells function properly, most Canadians consume far more than is necessary, or recommended. Most sodium is consumed as sodium chloride, also known as "table salt". Processed foods are the main source, accounting for 77% of average daily sodium intake. The maximum recommended daily salt intake from all sources (food, cooking, etc) is 1 tsp or 6 g.

Patient	Average Daily Salt Intake (g)	Diastolic Blood Pressure (mmHg)
A	8.4	80
B	10.4	105
C	4.9	78
D	7.8	87
E	12.8	112
F	11.9	108
G	8.4	96
H	11.4	88

Salt in our diet has been shown to contribute to high blood pressure. Blood

pressure (BP) is the pressure exerted by circulating blood upon the walls of blood vessels. During each heartbeat, BP varies between a maximum (systolic pressure), pressure in the arteries when the heart is contracting, and a minimum (diastolic pressure), pressure in the arteries when the heart is relaxing and expanding. A person's BP is expressed in terms of the systolic pressure over the diastolic pressure, measured in mmHg (mm of mercury, although mercury is not used), for example 120/80.

Draw a scatter plot on a set of axes, of the Diastolic BP vs. Salt Intake for the eight patients in the table.



a) Pose a question that would require one-variable data analysis.

b) Pose a question that would require two-variable data analysis.

c) Give three descriptive statements about this data.

3.2 Effective Surveys

Date: _____

Purpose: To learn how to design and conduct an effective and fair survey.

Definitions:

Primary Data _____

Secondary Data _____

Outlier _____

Influential Point _____

Population _____

Sample _____

Bias _____

2. **Warm-Up** (work in groups of 3):

a) Make a 3 – question survey to find out your classmates' musical preferences.

b) Do you think the participants can answer the questions honestly, without the questions having any influence on their answers?

c) Do any of the questions contain bias? If yes, how could these questions be rewritten so they are not biased?

3. As a class, discuss: How important is it to have accurate surveys and accurate results?

Refer to Summary of 10 Tips on Pg. 152

3 Principles of Proper Surveying (See Examples on Pg 154-155)

1. Ethics in Surveys:

- The introduction must include the intentions of the survey _____
- Respondents must be willing to participate.
- Keep it confidential. Do not ask for their names.
- Personal info must be kept private.

2. Design for Honest Responses:

- If answers are kept anonymous and confidential, participants are more likely to be honest with their responses.
- In **multiple choice questions**, include answers like "Do not know", "Not applicable", or "Other" as responses, so there is an option that is not part of a given list.
- Make sure **rating scales** are clear and offer a range of ratings (eg. On a scale of 1 to 10 how would you rate...)
- Do not have too many questions of the same type or require respondent to think hard about their responses.
- Ask more difficult or sensitive questions near end of survey and in a way that does not encourage the participant to lie.

3. Eliminate Bias:

- Avoid questions that lead to a certain answer or contain words that make the respondent think about negative images.
- Pose questions that are neutral to avoid bias that favours certain cultures or ways of life.
- Use a natural order of choices, such as "**Yes or No**" or "**Excellent, Very Good, Good, Fair...**" (from very positive to very negative).
- Consider mode of delivery (personal interview, by telephone, by mail, via the Internet) that is appropriate for the types of questions asked.

5 Types of Questions that can be used in a Survey

1. Dichotomous _____
2. Multiple Choice _____
3. Rating Scale _____
4. Completion _____
5. Open-ended _____

Homework: Pg 156 # 1 – 11

By the end of class: Write up the 5-6 survey questions for your Survey Assignment.

If there is time, have Mrs. Behnke check them over for bias and appropriateness for the assignment.

U3D3 **Survey Assignment is due:

It not be accepted after

In groups of 2-4, carry out a survey to investigate the relationship between two sets of numerical data. If you prefer to work alone, you may.

Design 4-6 survey questions without bias. You must use at least three of the five types of survey questions studied in class.

Define your population and describe how you selected your sample. Use a minimum of 25 participants. Organize the results in a table. Then draw a scatter plot of the data using excel. Make a conclusion regarding what you learned about this relationship.

Assignment will be marked for completion, organization, and neatness – see marking scheme on web-site before handing in your assignment.

By the end of today you should have

1. decided what numerical data you intend to collect and
2. created your survey questions.

Next class: carry out your survey, begin typing up your report

The final class: Type your relevant numerical data into excel and create a scatter plot.

Your assignment should be all on one piece of paper (print preview before printing).

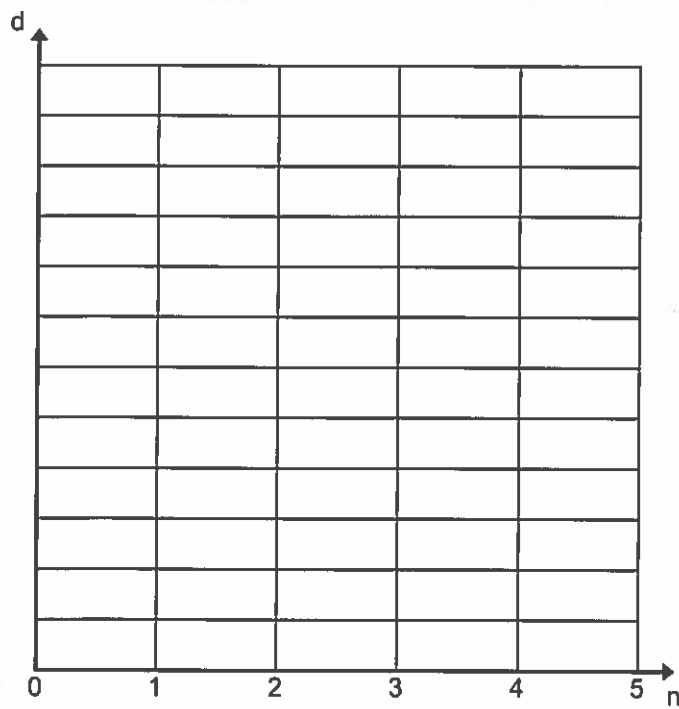
Examples of numerical data: How many hours you work per week. How much you earn on average per week. How many hours you sleep, on average, per week. How much you spend on Christmas gifts on average per year. How many hours you spend texting on average per day. How many hours you spend accessing the internet/cell service per day. How many times you buy your lunch per week. How many times you buy a beverage per week. How much you spend on clothes per year. How many times per year you get your hair cut. How much you spend on a haircut, on average. How much you spend on make-up per year. Your age. Your grade. Number of tv's in your home. Number of computers in your home (including laptops but not tablets). How many tablets and phones in your home. Number of people living in your home. Number of pets you have had so far in your lifetime, Your weight, your maximum bench press, your maximum dead lift Etc., etc.

Participant			
A			
B			
C			
D			
E			
F			
G			
H			
I			
J			
K			
L			
M			
N			
O			
P			
Q			
R			
S			
T			
U			
V			
W			
X			
Y			

1. Complete Investigation Pg 160 # 1 – 6

Accuracy

Name	Toss 1	Toss 2	Toss 3	Toss 4	Toss 5



Answer this question: “Does practice make perfect in a statistical sense?”

U3D4: Answer Pg 165 # 1 – 3 in your notes

NEXT DAY: U3D5 GO TO COMPUTER LAB ... check wodss website MRBS at bottom to see which lab.

Finish collecting the data for your survey assignment. (Survey at least 25 people) then use computer to finish assignment. Survey assignment is due at the end of next class

U3D5 Finish Survey Assignment on Computer in room Using Microsoft Excel

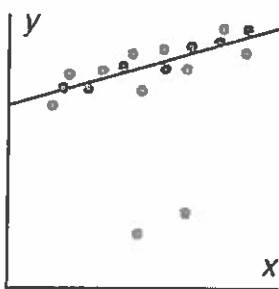
***Survey assignment is due at the end of today’s class. (Check Mrs. Behnke’s web-site for marking scheme.)**

Line of Best Fit:

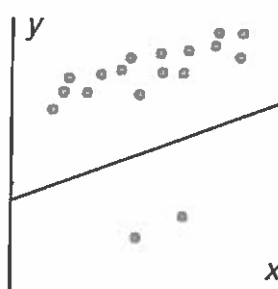
Drawing a line of best fit: Find a pathway through the middle of the data. The more spread out the data is the more difficult it is to draw the line of best fit. The line of best fit should reflect all valid points including outliers.

Example 1) Which line is the line of best fit? Justify your choice.

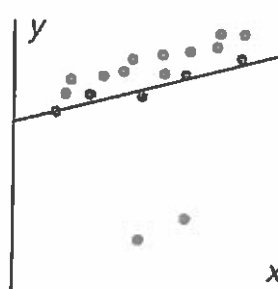
Graph A



Graph B



Graph C



A line of best fit can be used to make predictions.

Interpolation: _____

Extrapolation: _____

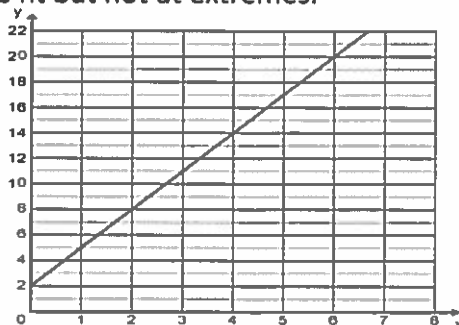
Data Spread and Reliability How confident can we be of predictions made from scatter plots?

A model with data spread over a larger interval is more reliable than data spread over a smaller interval. The farther we get from the main cluster the less confidence we can have on the predictions we make.

Sample Size and Reliability The more data we use, the more reliable the prediction should be.

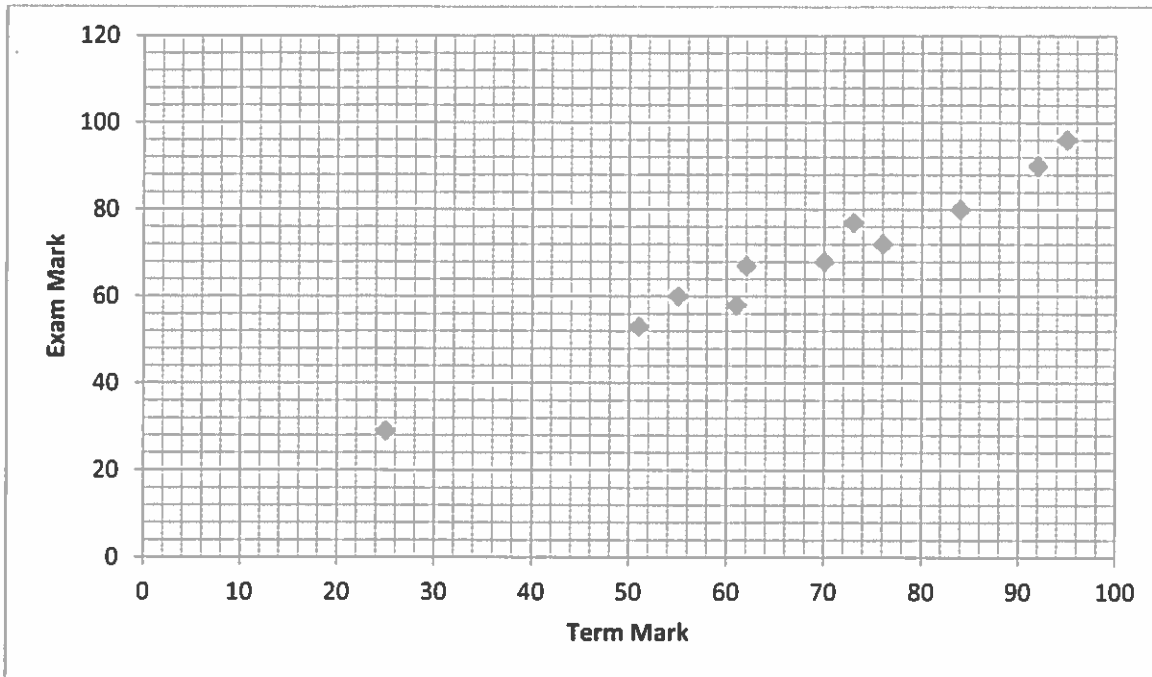
Non Linear Data Not all relationships between variables are linear. Over a small interval a linear model may be a reasonable fit but not at extremes.

Example 2)



Find the equation of the line of best fit.

Example 3) These are pre-exam term marks and exam marks for some students in a gr. 12 English course.



- a) Draw a line of best fit
- b) Determine an equation of the line of best fit.

- c) Use the data to predict the exam mark of a student with a term mark of 98%. Is this Interpolation or Extrapolation?

- d) Use the data to predict the exam mark of a term mark of 10%. Is this Interpolation or Extrapolation?

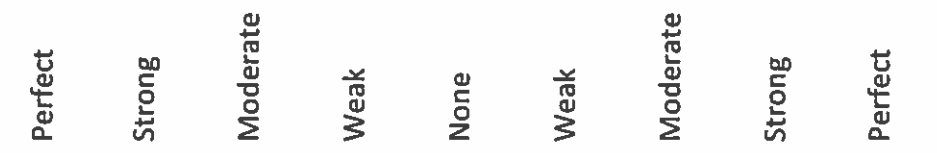
Linear Correlation

Linear Regression

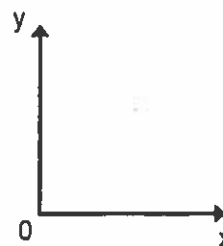
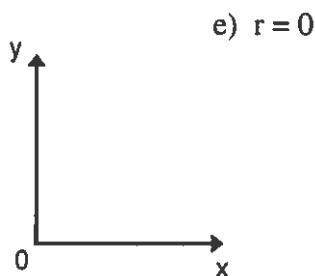
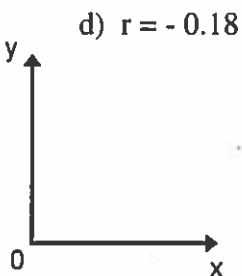
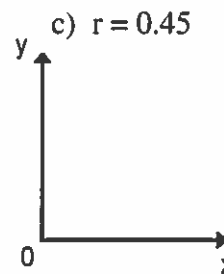
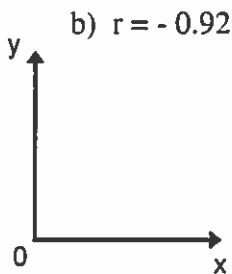
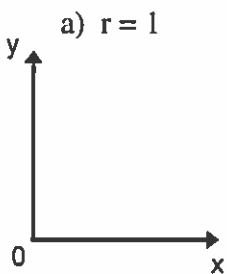
Through a series of calculations, a **value r , which is called the correlation coefficient, can be assigned to represent how far the points are, on average from the line of best fit.**

The closer the value of r is to $+1$ or -1 , the closer the fit. The sign of r relates to the slope of the line, not the fit of the line. r does not indicate whether there is a direct relationship between the variables.

The following diagram illustrates how the correlation coefficient r corresponds to the strength of a linear correlation.



Ex. 1 Sketch the general appearance of a scatter plot associated with each value of the correlation coefficient r . Include a line of best fit where applicable.



Ex. 2 Regression Analysis

Once you determine that two variables have a moderate to strong linear relationship, you can determine a linear model. This will allow you to make predictions for one variable based on the value of the other variable. Linear regression software, eg. the TI83+ graphing calculator can perform many complex calculations to determine the equation of the line of best fit between two variables.

Student	Aptitude Test Score	Final Math Mark
A	95	86
B	73	68
C	59	61
D	68	70
E	84	90
F	80	87
G	82	71
H	64	60
I	66	72
J	50	45
K	74	74
L	64	53

Last year, twelve randomly selected students took a mathematics aptitude test before they began their college mathematics course. Their results on the aptitude test and their final mark in the course are shown in the table. All scores are out of 100.

- Make a scatter plot of the data and use linear regression to determine the equation of the line of best fit. Give the slope and vertical-intercept.
 - How well does the regression equation fit the data? Explain.
- c) If Hannah scored 75 on the aptitude test, what would you expect her mark to be in her college math course? (On TI-83 plus – use 2nd Trace to calculate... enter 75 as x-value in Desmos, click on the line and slide cursor until you see an x-value of 75)
- d) If Pietro scored 40 on the aptitude test, what would you expect his mark to be in his college math course?

Method to answer part (a): Google Desmos
www.Desmos.com/calculator
*** Desmos Instructions

Click on + sign

Choose table, enter data

Zoom out to see data on graph

Click on + sign again

Choose Expression

enter $y \sim mx + b$

Observe r , r^2 , m and b

Cause and Effect Relationship _____

Eg. _____

1. For each independent variable, identify a dependent variable that might form a cause and effect relationship.

- a) The amount of time students study for an exam.

- b) The cost of gasoline sold.

- c) The amount of space used to display a product in a store.

- d) The amount of time a person exercises per week.

- e) The average number of cars driven in a city per day.

Regression Analysis _____

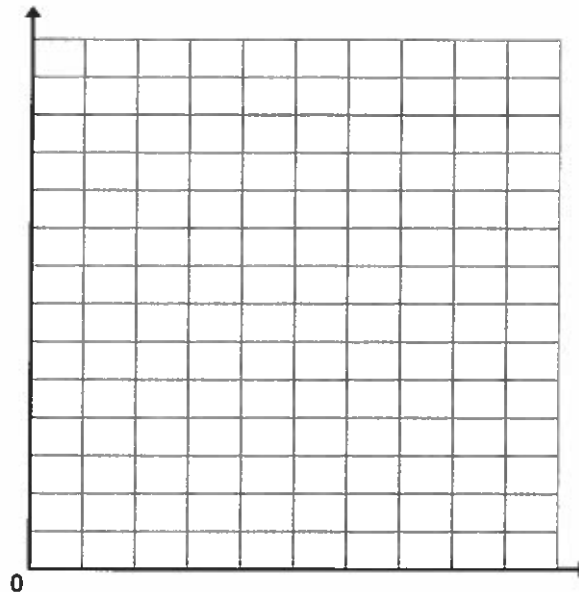
Errors in Analysis occur when:

- 1. There is too little data.
- 2. Using linear regression (line of best fit) for a non-linear relation.
- 3. Using linear regression (line of best fit) when the correlation is weak.
- 4. Reversing the cause and effect relationship.
- 5. Extrapolating outside the range of the data set.
- 6. Not considering the effects of outliers or influential points.

A high correlation for a data set does not always indicate a cause and effect relationship between two variables. Often, more data and analysis are needed to prove such a relationship exists.

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

Acres of Land	Number of 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



- Which variable is the independent variable?
- Which variable is the dependent variable?
- Create a scatter plot. Choose the appropriate axis for each variable. (i.e. remember which axis is used for the independent variable and which is used for the dependent variable). Label each axis, choose an appropriate scale for each axis and give your graph a title.
- Circle any outliers or influential points.
- Describe the trend you see. (What is the correlation?)
- Describe the relationship between the variables.
- Draw the line of best fit.
- Determine the equation of the line of best fit.
- How many parking spaces would you expect on 10 acres of land? Use your equation to solve. Is this interpolation or extrapolation?
- How many parking spaces would you expect to fit on 3.5 acres of land? Use your graph to solve. Is this interpolation or extrapolation?