Unit 5 – Statistics (Data Analysis)

Introduction to Statistics

Understanding Statistics Requires a Change in Mindset

Statistics: A branch of mathematics concerned with data collection, presentation, analysis and interpretation.

Mindset: A way of thinking; an attitude or opinion, especially a habitual one.
e.g. *Earth Day is a way of propagating and celebrating the environmentalist* ***mindset****.*

In many ways, this unit will remind you of the previous two units. This unit deals with

- Relations (mathematical relationships), especially linear relations
- Graphs
- Equations, especially equations of lines in slope-*y*-intercept form
- Slopes and *y*-intercepts
- Rates of change and initial values
- Substituting values into equations and solving for the unknown

However, the ***focus*** of this unit is entirely different. The above concepts are used to help us understand ***real-world*** data! This unit ***mainly*** deals with

- Collecting data involving two variables
- Creating scatter plots
- Creating lines of best fit (estimating using pencil and paper as well as with software such as TI-Interactive)
- Understanding and evaluating the "strength" of the relationship between the two variables (i.e. the ***correlation***)
- Accepting that real-world data rarely if ever allow for a "perfect fit" to an equation
- Finding equations of lines of best fit
- Using equations of lines of best fit to predict values of variables
- Evaluating how good our estimates are
- Estimating by interpolating and extrapolating

Understanding Scatter Plots #1

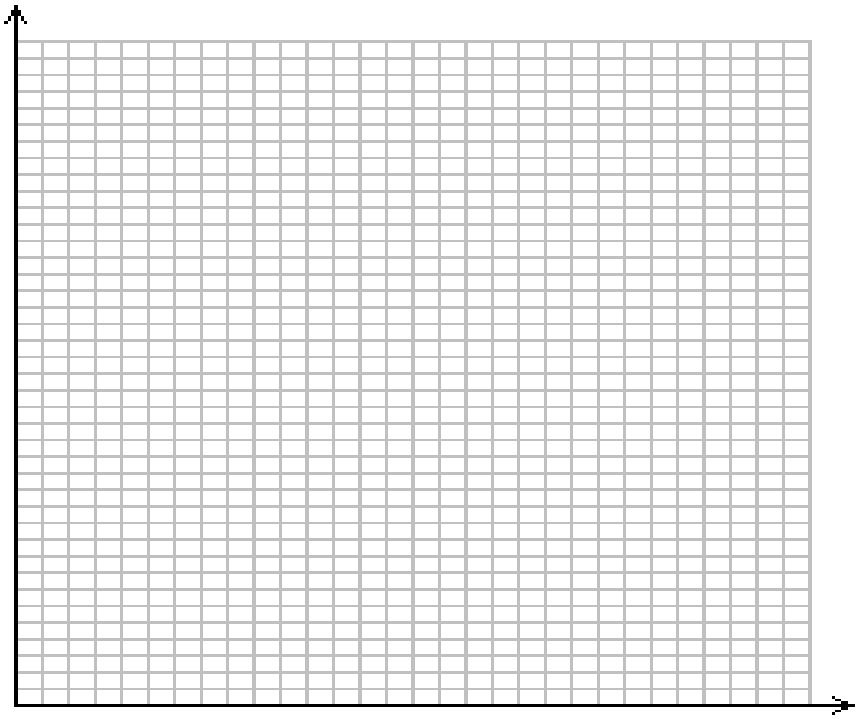
1. Eliseo performed a study to explore how TV viewing habits affect student performance. He collected data by surveying several students in his math class. He asked each student to provide their current math mark as well as the number of hours spent watching TV each day. The data are summarized in the following table:

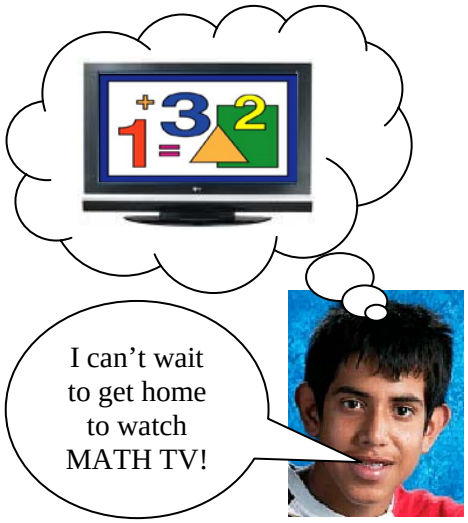
Daily TV Viewing Time (h)	1	6	3	2	10	0	2	5	2	12	8	5	2	4
Mark (%)	83	53	71	73	81	95	68	51	70	40	21	32	75	27

(a) State the independent and dependent variable.

Independent: _____ **Dependent:** _____

(b) Create a ***scatter plot*** of the data (plot the data points). ***Do not connect the dots!*** Label the axes and include a title for your graph. In addition, ***circle any outliers***.





(c) Describe the relationship between the students' daily TV viewing time and their mathematics marks.

(d) Draw a line of best fit. Then write the slope-*y*-intercept equation for the line of best fit. Show your work!

Equation of Line of Best Fit: _____

(e) Use the equation of your line of best fit to estimate the math mark of a student who watches four hours of TV per day.

(f) Again using your equation, estimate the number of hours of TV watched by a student with a mark of 45%.

(g) How certain are you that your estimates are accurate?

(h) Now check your answers to (e) and (f) by using your graph.

Equation Answer	***Graph Answer***	***Do the answers agree?***
(e)	(e)	
(f)	(f)	

2. Now use TI-Interactive to create a scatter plot and to determine the line of best fit for the same data given in question 1. Print out the TI-Interactive document that you create and staple it to this sheet. In addition, summarize your results below.

Equation obtained using your line of best fit: _____

Equation obtained using TI-Interactive: _____

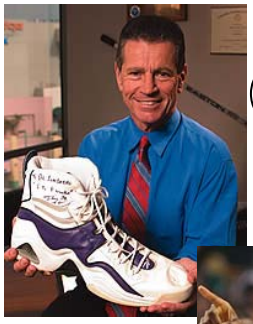
3. Complete the following table. Use point form.

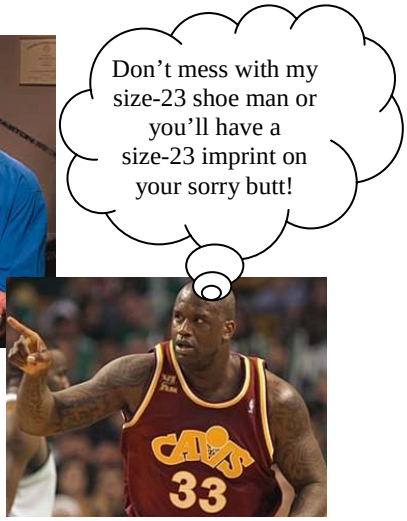
Similarities between Unit Four and Unit Five	***Differences between Unit Four and Unit Five***

Understanding Scatter Plots #2

Predicting Shaquille O'Neal's Hand Span

In this activity you will collect data by measuring foot lengths and hand spans. You will then use your data to predict Shaquille O'Neal's hand span.





Step One – Calculating Shaq's Foot Length in Centimetres

It is well known that Shaquille O'Neal (also known as "Shaq") wears a size-23 shoe. What is not well known is his foot length in centimetres. Luckily, there are formulas that relate shoe size, as measured with a Brannock device (see diagram below and to the right), to foot length, in inches.

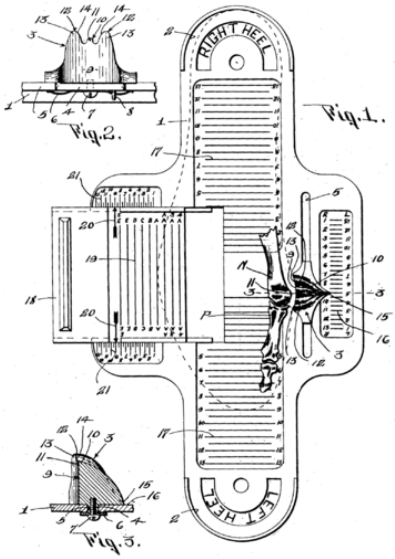
$m \rightarrow$ represents men's shoe size as measured by a Brannock device

$w \rightarrow$ represents women's shoe size as measured by a Brannock device

$f \rightarrow$ represents foot length ***in inches***

$$\boxed{m = 3f - 22} \qquad \boxed{w = 3f - 21}$$

(a) Use the appropriate formula above to calculate Shaq's foot length ***in inches***. Show all work!



A Brannock Device

(b) Now convert Shaq's foot length ***to centimetres*** by using the equation $C = 2.54I$, where C represents the length in centimetres and I represents the length in inches.

Conclusion: Shaq's foot length in cm is _____ .

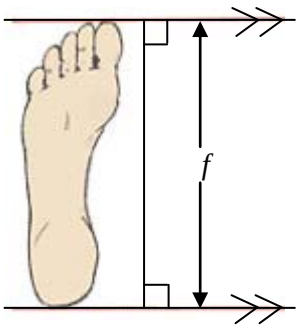
Step Two – Collecting the Data by Measuring Hand Span and Foot Length

How to Measure Hand Span

- The hand is placed ***palm down*** on a flat surface.
- The fingers are outstretched as far as possible.
- Measure the distance between the ***outside of the thumb*** to the ***outside of the little finger***.

How to Measure Foot Length

- Shoes must be removed.
- Place the most prominent toe and the most prominent part of the heel between two parallel lines that are perpendicular to the foot.
- Measure the distance between the two parallel lines.



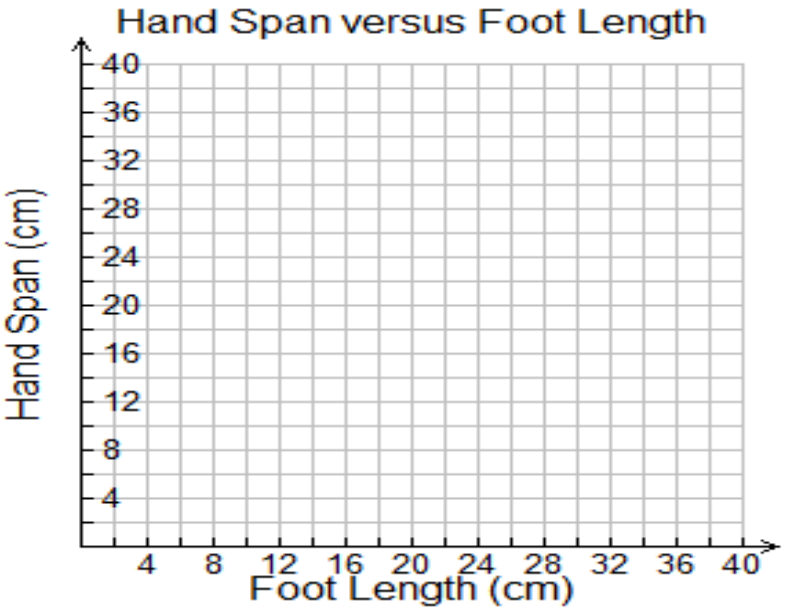
Use the measuring procedures described above to complete the following table.

Measure to the nearest millimetre, that is, to one decimal place.

Student Name	***Foot Length in cm (f)***	***Hand Span in cm (h)***	***Ratio of $f:h$***
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			

Step Three – Analyzing the Data

(a) Use the data in the table to create a scatter plot. ***Do not connect the dots!***



(b) Carefully construct a line of best fit. Determine its equation, its slope and its intercepts.

Equation: _____ Slope: _____ Vertical Intercept: _____ Horizontal Intercept: _____

(c) Now use TI-Interactive, a graphing calculator or spreadsheet software to determine the equation of the line of best fit as well as its intercepts. Summarize your results below.

Equation: _____ Slope: _____ Vertical Intercept: _____ Horizontal Intercept: _____

(d) Explain why it is better to use the equation obtained in (c) than it is to use the equation obtained in (b).

(e) Explain the ***meaning***, in the context of this problem, of each of the following.

Slope = Constant of Variation = Rate of Change: _____

Vertical Intercept = Initial Value: _____

Horizontal Intercept: _____

(f) Does the data that you collected show a positive correlation, a negative correlation or no correlation? Explain.

(g) If you did everything correctly, your line of best fit should have a positive slope. Explain why you would expect this.

Step Four – Predicting Shaq's Hand Span

(a) You will use two different methods to predict Shaq's hand span.

Method 1	***Method 2***
Use the equation from (c) in step 3.	Calculate the average of the $f:h$ ratios from the table on the previous page. Then use this average to predict Shaq's hand span.
Using method 1, I predict Shaq's hand span to be:	Using method 2, I predict Shaq's hand span to be:
_____	_____

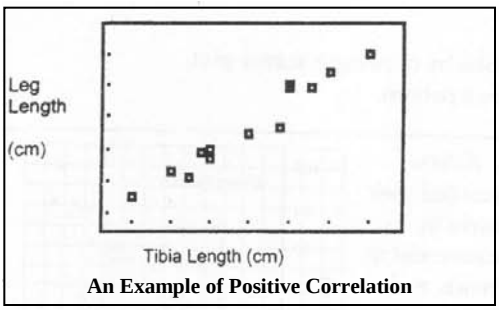
(b) Predicting Shaq's hand span is an example of ***interpolation / extrapolation*** **(circle the correct answer)** because

_____ .

(c) The correct answer to the previous question is "extrapolation." In the space provided below, show an example of ***interpolation*** that involves the data you collected in step 3.

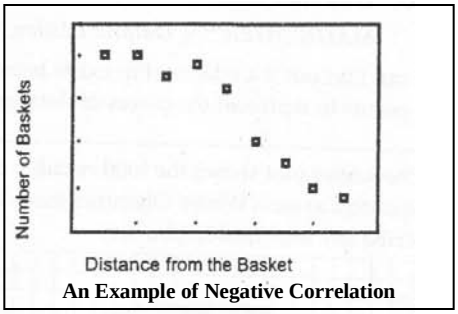
Scatter Plots and what they Tell us

Investigating Correlation



An Example of Positive Correlation

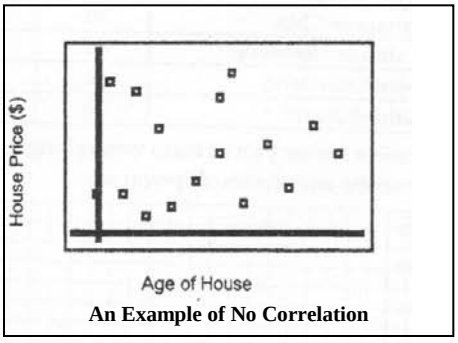
1. The graph shows the plotted points rising upwards to the right.
   - Agree
   - Disagree
   - Pass
2. As the length of the tibia increases the length of the leg increases
   - Agree
   - Disagree
   - Pass
3. The graph can be used to determine the length of a person's leg if you know the length of the tibia bone
   - Agree
   - Disagree
   - Pass



An Example of Negative Correlation

1. The graph shows the plotted points falling to the right
   - Agree
   - Disagree
   - Pass
2. As the distance from the net increases the number of baskets made decreases.
   - Agree
   - Disagree
   - Pass
3. The graph can be used to determine the number of baskets you will make if you know the distance from the basket.
   - Agree
   - Disagree
   - Pass

1. The graph shows the plotted points scattered.
   - Agree
   - Disagree
   - Pass
2. As the age of the house increases the price of the house is either large or small.
   - Agree
   - Disagree
   - Pass
3. The graph can't be used to determine the price of the house if you know how old it is.
   - Agree
   - Disagree
   - Pass



An Example of No Correlation

What is Correlation?

In ***statistics***, the ***correlation coefficient*** is used to measure the ***"strength" of the relationship between two variables***. Researchers collect data (make measurements of some kind, usually involving two variables) and then try to determine whether the variables are related to each other. The purpose of this process is to help us make predictions about one variable based on what we know about another variable.

For example, there is a correlation between income and education. We find that people with higher income usually have more years of education. When we know there is a correlation between two variables, we can make a prediction. If we know a group's income, we can predict their years of education.

Direction

There are two ***types*** or ***directions*** of ***correlation***. In other words, there are two patterns that correlations can follow. These are called ***positive correlation*** and ***negative correlation***.

Positive Correlation

In a positive correlation, as the values of the independent variable increase, the values of the dependent variable also tend to increase. The example above of income and education is a positive correlation. People with higher incomes also tend to have more years of education. People with fewer years of education tend to have lower income.

Here are some examples of positive correlations:

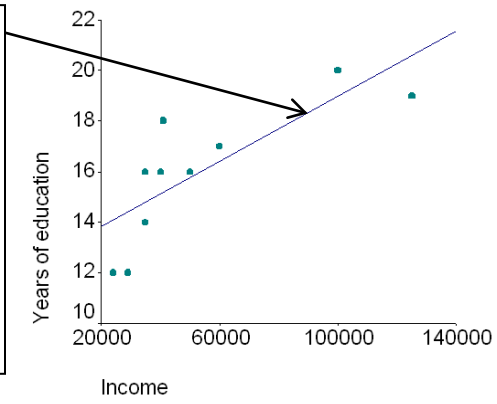
1. SAT scores and college achievement—among college students, those with higher SAT scores also have higher grades
2. Happiness and helpfulness—as people's happiness level increases, so does their helpfulness

This table shows some sample data. Each person reported income and years of education.

Participant	*Income*	*Years of Education*
#1	125,000	19
#2	100,000	20
#3	40,000	16
#4	35,000	16
#5	41,000	18
#6	29,000	12
#7	35,000	14
#8	24,000	12
#9	50,000	16
#10	60,000	17

Line of Best Fit

Notice that the line of best fit is drawn as close to the data points as possible. There should be about as many points above the line as there are below the line.



We can make a graph, which is called a scatter plot. On the scatter plot, each point represents one person's answers to questions about income and education. The line is the best fit to those points. All positive correlations have a scatter plot that looks like this. The line will always go in that direction if the correlation is positive.

Negative Correlation

In a negative correlation, as the value of the independent variable increases, the value of the dependent variable decreases. The word "negative" is a label that shows the direction of the correlation.

Here are some other examples of negative correlations:

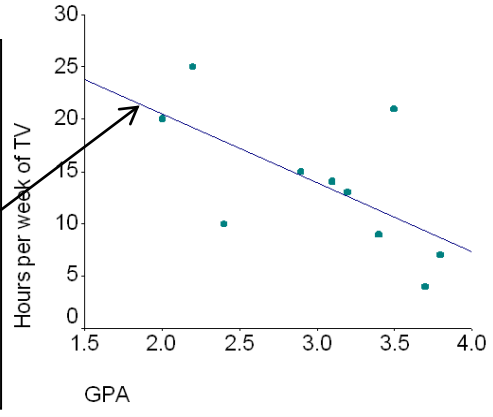
1. Education and years in jail—people who have more years of education tend to have fewer years in jail (or phrased as people with more years in jail tend to have fewer years of education)
2. Crying and being held—among babies, those who are held more tend to cry less (or phrased as babies who are held less tend to cry more)

The scatter plot below shows the sample data from the table. The line on the scatter plot shows what a negative correlation looks like. Any negative correlation will have a line with that direction.

Participant	*GPA*	*TV in hours per week*
#1	3.1	14
#2	2.4	10
#3	2.0	20
#4	3.8	7
#5	2.2	25
#6	3.4	9
#7	2.9	15
#8	3.2	13
#9	3.7	4
#10	3.5	21

Line of Best Fit

Notice that the line of best fit is drawn as close to the data points as possible. There should be about as many points above the line as there are below the line.

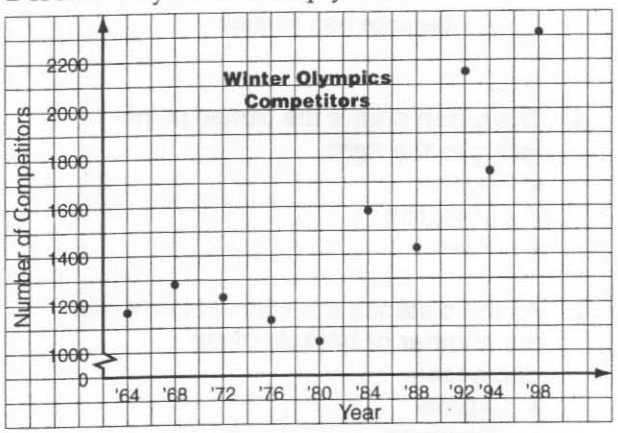


4.10 Scatter Plots

MATHPOWER™ 9, Ontario Edition, pp. 204–208

You can find out if a relationship exists between two variables by drawing a **scatter plot**. Plot points to represent the pieces of data, and then look for a pattern.

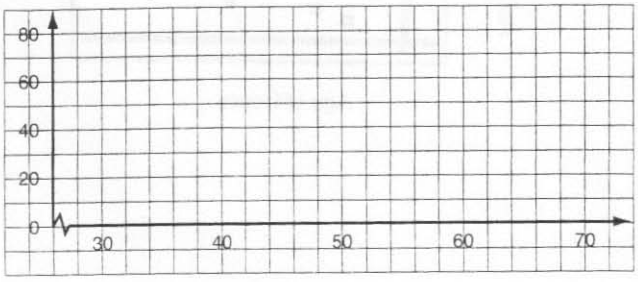
1. The scatter plot shows the total number of competitors at each Winter Olympics since 1964. Describe any relationship you see.



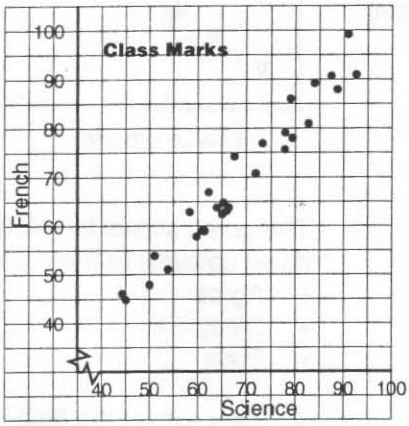
2. The table gives average heights and average masses of different types of dogs.

Type of Dog	Height (cm)	Mass (kg)
Belgian Sheepdog	63	30
Cocker Spaniel	42	14
Collie	63	27
English Springer Spaniel	51	23
Irish Setter	69	32
Irish Terrier	48	12
Japanese Chin	30	4
Labrador Retriever	60	30
Newfoundland	71	65
Saint Bernard	70	57

Draw a scatter plot of mass versus length. Describe any relationship you see.



3. A class recorded their marks in science and in French, and drew a scatter plot of the data.

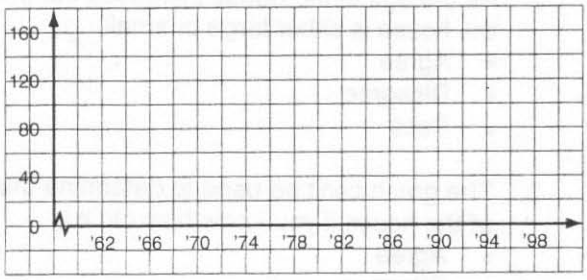


From this data, does it appear that a student's science mark is a good indication of the French mark? Explain.

4. The table gives the total population of whooping cranes in North America over several years.

Year	Population	Year	Population
1962	45	1978	75
1964	35	1980	77
1966	40	1982	73
1968	50	1984	82
1970	60	1986	110
1972	52	1988	130
1976	68	1990	140

a) Draw a scatter plot of population versus year.



b) Use your scatter plot to estimate the total whooping crane population in 1994 and 1998.

c) Use your research skills to find the actual numbers in 1994 and 1998. Compare with your estimates.

4.11 Lines of Best Fit

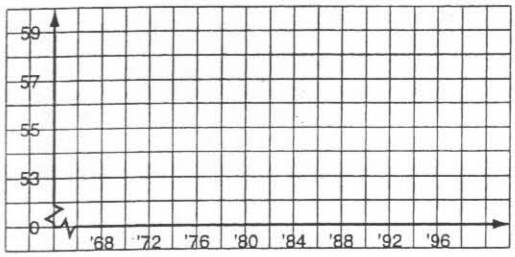
MATHPOWER™ 9, Ontario Edition, pp. 209–211

The **line of best fit** on a scatter plot is a line drawn as close as possible to all the data points. There are about as many points above the line as there are below the line.

1. The table gives the winning times in the men's 100-m backstroke swimming event at several Summer Olympics.

Year	Winner	Winning Time(s)
1968	Matthes (E. Germany)	58.7
1972	Matthes (E. Germany)	56.58
1976	Naber (U.S.)	55.49
1980	Baron (Sweden)	56.33
1984	Carey (U.S.)	55.79
1988	Suzuki (Japan)	55.05
1992	Tewksbury (Canada)	53.98

a) Draw a scatter plot of winning time versus year. Draw a line of best fit.



b) In 1960, David Thiele of Australia won the event. Extrapolate to estimate his winning time.

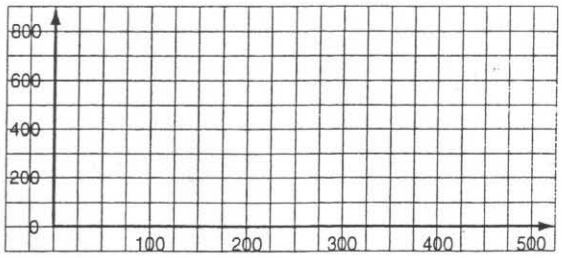
c) David Thiele's actual winning time was 61.0 s. Compare this time with your estimate. How close were you?

d) Estimate the winning time in 2000.

2. The table gives the winning times for five women's freestyle swimming events at the 1992 Summer Olympics.

Distance	Winning Time(s)
50 m	24.79
100 m	54.64
200 m	117.90
400 m	247.18
800 m	505.52

a) Draw a scatter plot of distance versus winning time. Draw a line of best fit.



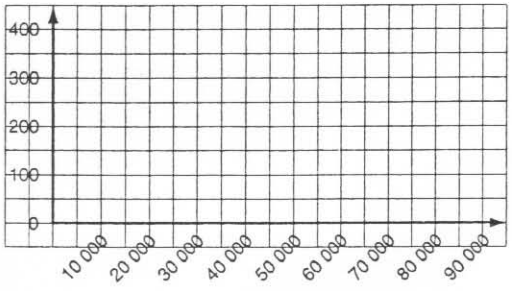
b) If there had been a 600-m race, what winning time would you estimate for it?

c) If there had been a 1000-m race, what winning time would you estimate for it?

3. The table gives the areas and maximum depths of several Canadian lakes.

Lake	Area (km^2)	Maximum Depth (m)
Superior	83 270	393
Huron	60 700	229
Michigan	58 020	281
Great Bear	31 790	319
Great Slave	28 440	140

a) Draw a scatter plot of maximum depth versus area. Draw a line of best fit.



b) The area of Lake Erie is 25 680 km^2. Use your scatter plot to predict the maximum depth of Lake Erie.

c) The actual maximum depth of Lake Erie is 64 m. Compare this with your estimated depth. Does your line of best fit give reasonable estimates of maximum lake depths? Explain.

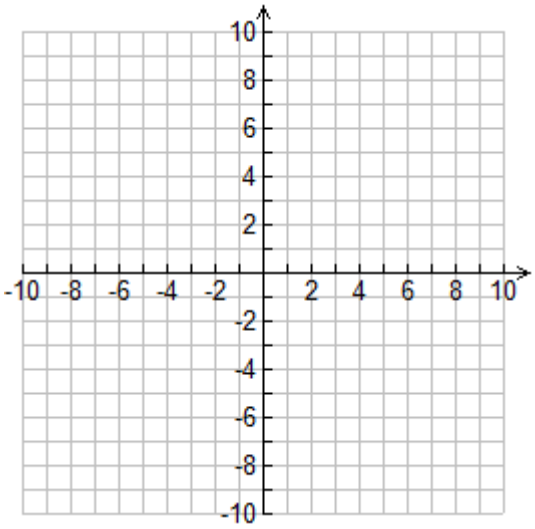
Linear and Non-Linear Relations Revisited

1. Given the table of values, identify whether each relation is linear or non-linear. **Explain** your reasoning.

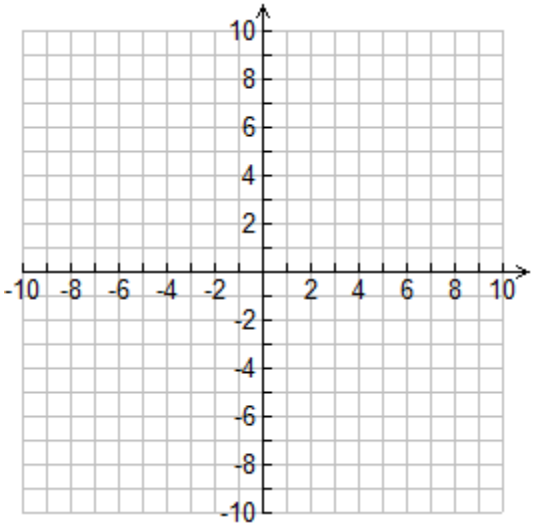
x	*y*	first differences
1	–1	
2	1	
3	3	
4	5	
5	7	

x	*y*	first differences
–2	1	
–1	–2	
0	–3	
1	–2	
2	1	

2. Graph each of the relations in question 1. Determine an equation for each relation.



Equation: _____



Equation: _____

3. For each graph shown on the grid, state whether the graph represents a linear or non-linear relation.

Linear or non-linear?

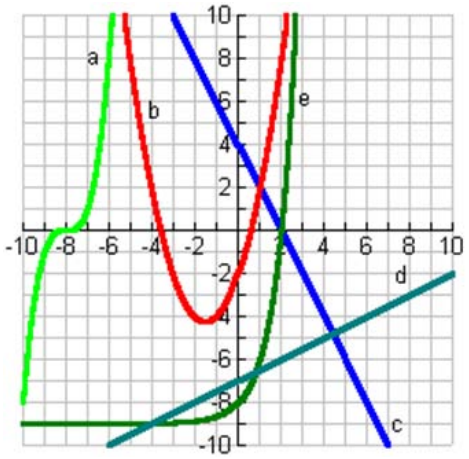
(a) _____

(b) _____

(c) _____

(d) _____

(e) _____



4.

Dependent and Independent Variables

A change in the independent variable *causes* a change in the dependent variable. For example, a change in the number of bus passengers (independent variable) affects the weight of the bus (dependent variable).

1. For each pair of quantities, decide which is the independent variable and which is the dependent variable. Draw an arrow from the independent variable to the dependent variable. For example,

bus passengers ⟶ weight of bus.

a) number of customers — total sales

b) body temperature — time spent in cold shower

c) average traffic speed — number of cars on the highway

d) number of schools in a city — total population of city

e) number of cigarettes smoked per day — money saved

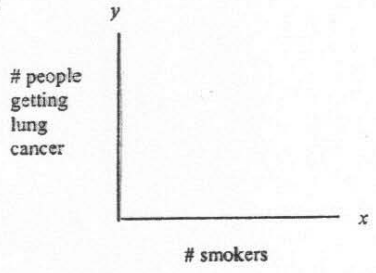
f) number of traffic accidents — number of drunk drivers

g) humidity level — sales of air conditioners

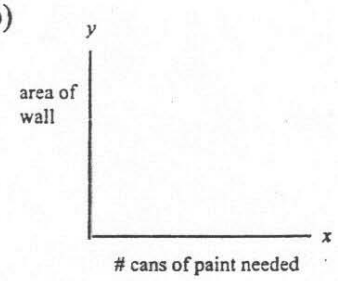
h) number of homes flooded — amount of rainfall

2. The independent variable should always be on the x-axis. Using this rule, label each set of axes as "correct" or "incorrect".

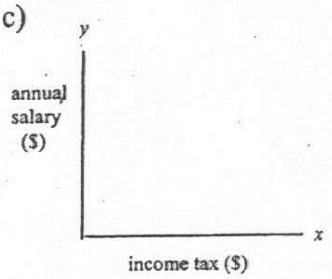
a)



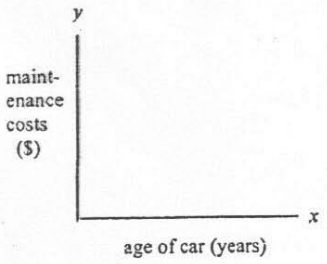
b)



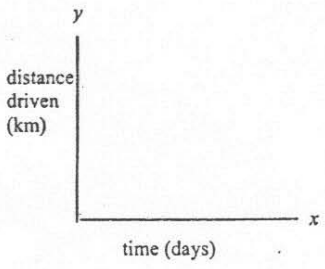
c)



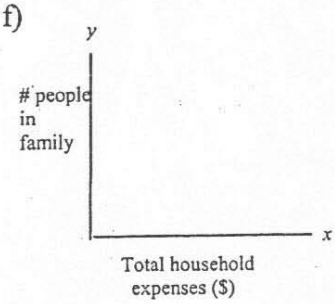
d)



e)



f)

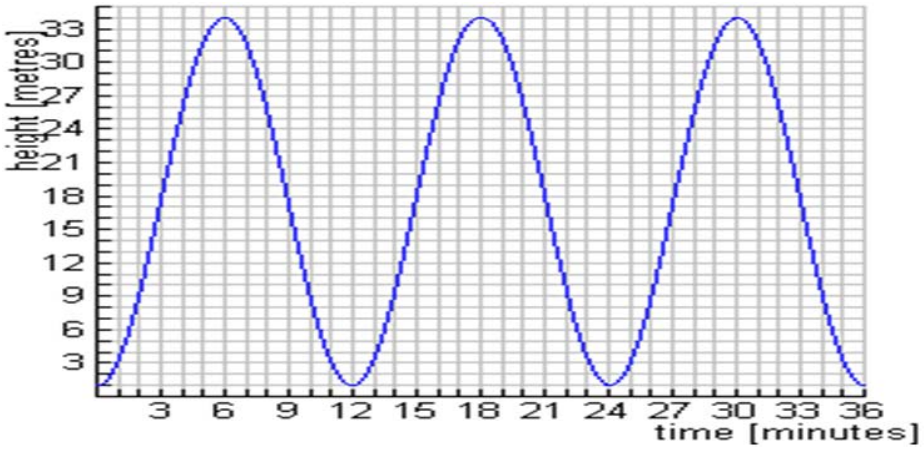


5. Consider the graph of the relation shown below.

(a) Identify the independent and dependent variables.

independent: _____ dependent: _____

(b) Describe how the dependent variable changes as the independent variable increases.



(c) Use the graph to estimate the height at 9 minutes. _____

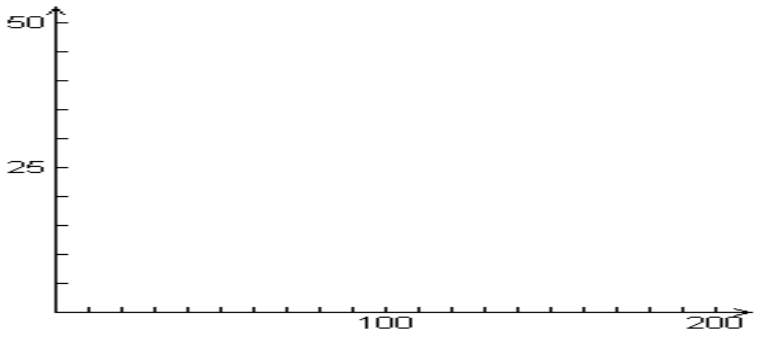
(d) Use the graph to estimate the times at which the height is 27 metres. _____

(e) This relation is said to have a ***periodic*** behaviour. Give at least one real-life example of what this relation could model.

6. High Energy Gas Company charges its customers $12 per month plus ten cents per cubic metre of gas used. The New Gas Company charges $20 a month plus five cents per cubic metre.

(a) For each company, write an equation to represent the total cost per month (C) in terms of the number of cubic metres of gas used (n).

(b) Identify the independent and dependent variables.



(c) Graph the relationship for each company**.**

(d) Use both the graphs and the equations to determine the following:

(i) the cost for a usage of 80 cubic metres of gas

(ii) the usage of gas for a cost of $50.00

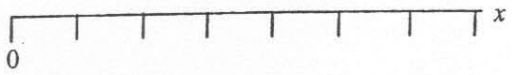
7.

Name: _____

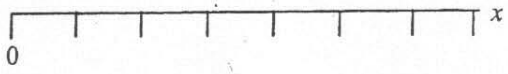
Choosing Scales for a Scatter Plot

1. Create a scale on the x-axis for each of the following situations, so that the distance from the lowest x-value to the highest x-value covers at least half the length of the axis. Use a break if necessary.

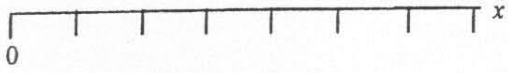
a) lowest x-value = 22
highest x-value = 130



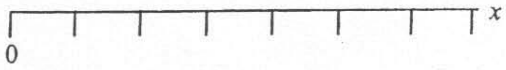
b) lowest x-value = 165
highest x-value = 511



c) lowest x-value = 82
highest x-value = 103



d) lowest x-value = 0.7
highest x-value = 2.8

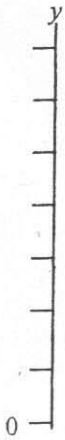


2. Create a scale on the y-axis for each of the following situations, so that the distance from the lowest y-value to the highest y-value covers at least half the length of the axis. Use a break if necessary.

a) lowest y-value = 3,200
highest y-value = 5,700

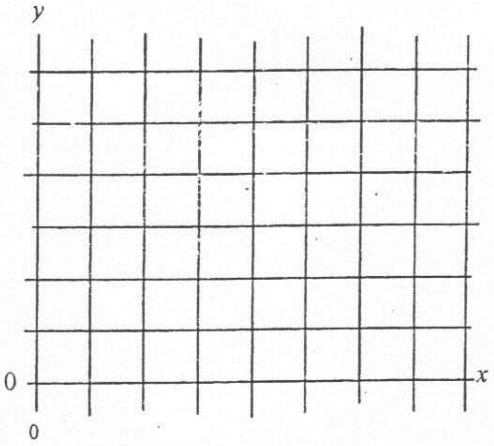
b) lowest y-value = 40.25
highest y-value = 41.3

c) lowest y-value = 30,000
highest y-value = 125,000



3. Create a scatter plot of the following data, showing the number of customers and total sales for several gas stations on the same day. Put the number of customers on the x-axis and sales on the y-axis and choose appropriate scales. Label the axes and put a title at the top.

station	#customers	Sales ($)
Gerrard St.	460	9,605
Woodbine Ave.	501	11,022
Main St.	455	8,645
Kingston Rd.	658	17,108
Jones Ave.	524	13,672
Pape Ave.	620	15,511
Mortimer Ave.	589	12,958
Sammon Ave.	607	11,836
O'Connor Ave.	570	11,970
Leslie St.	695	16,070

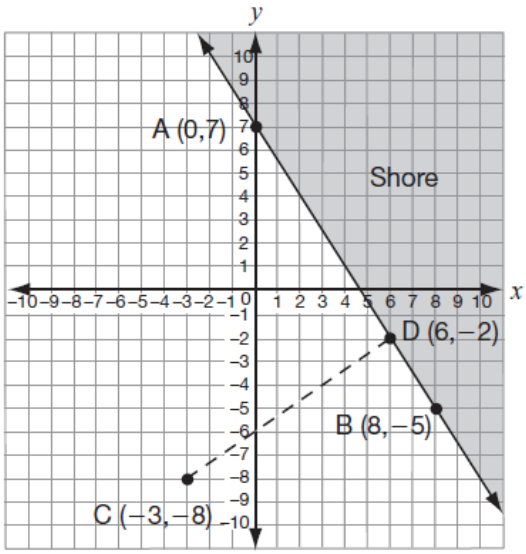


EQAO Practice – Linear Relations and Data Analysis

1. *Analytic Geometry and Linear Relations*

Washed Up on the Shore

A boat is travelling from Point C toward Point D, which is on the shoreline.
The shoreline is represented by the line through points A and B.



Determine whether the path from C to D is perpendicular to the shoreline.

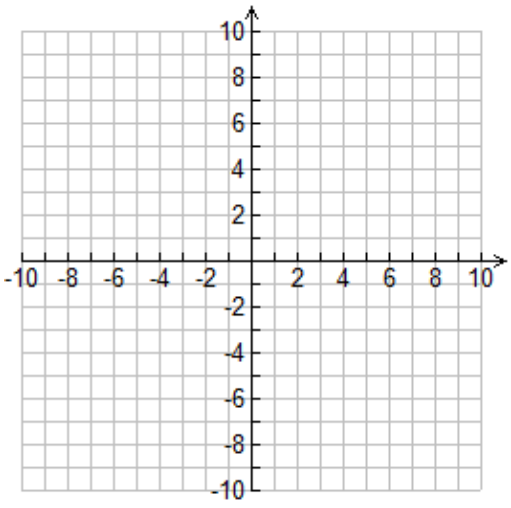
Justify your answer.

Excellent Equations

A line is perpendicular to the line $y = 2x + 3$ and has the same ***x*-intercept** as $x + 3y + 10 = 0$.

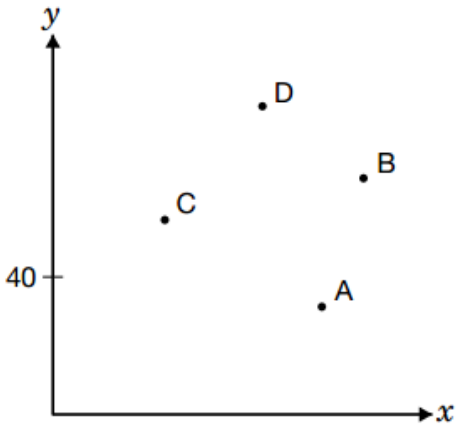
Find the equation of this line. Express your answer in the form $y = mx + b$.

Justify your answer.



Lineup

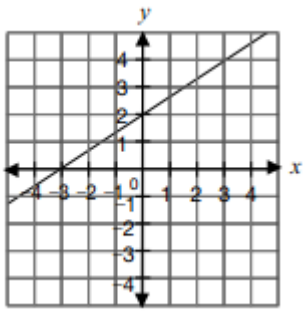
The line $y = \frac{1}{5}x + 50$ passes through only one pair of points below.



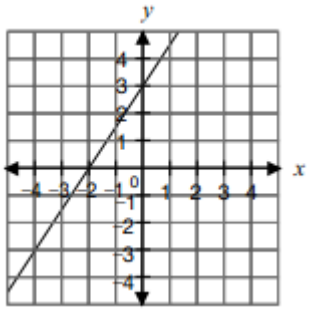
Which pair of points could the line pass through? Justify your response.

14 Which **graph** represents the relation $y = \frac{2}{3}x + 2$?

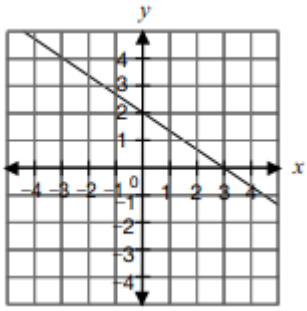
F



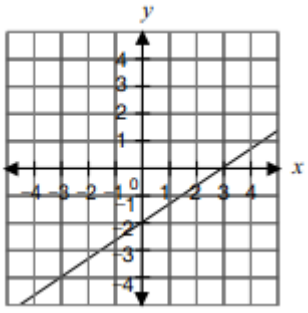
G



H



J



15 A line has the following characteristics.

- It is perpendicular to the line $y = \frac{1}{2}x + 3$.
- It passes through the point (4, 0).

What are m, the slope, and b, the y-intercept, of the line?

A $m = \frac{1}{2}; b = 0$

B $m = \frac{1}{2}; b = 3$

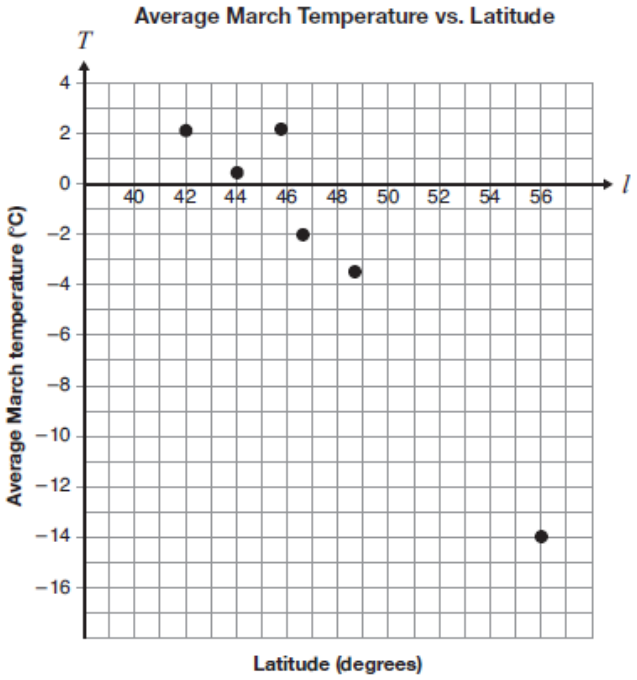
C $m = -2; b = 0$

D $m = -2; b = 8$

2. *Data Analysis – Scatter Plots and Lines of Best Fit*

March Temperatures

The average March temperatures for six Ontario communities are plotted according to their latitudes on the following scatter plot.



The city of Kenora has a latitude of 50° and has an average March temperature of −6.3 °C. Does the community of Kenora follow the trend of the data?

Justify your answer.

Wing Length

Wing length is a reliable method for determining the age of young birds. Below is an example of data for a particular species.

Wing length (cm)	Age (days)
1.5	4
3.1	8
3.2	10
4.1	12
5.2	16

Determine the age of a bird with a wing length of 3.6 cm.

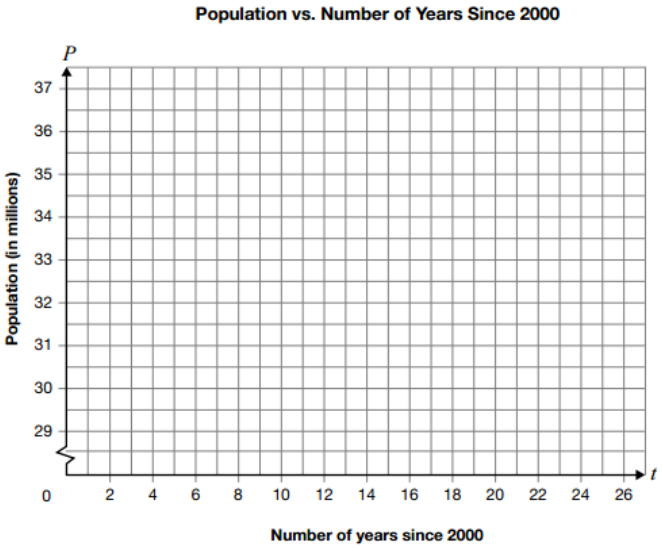
You may use the grid if you wish.

Justify your answer.

Population Plans

Alvin is researching the population of Canada. He finds data for the year 2001 and predictions for every 5 years after that, as shown below.

Number of years since 2000, t	Population (in millions), P
1	31.1
6	32.2
11	33.4
16	34.4
21	35.4
26	36.2



Determine an algebraic model for Alvin's data, and use it to make a reasonable prediction for the population of Canada in 2036.

Justify your answer.

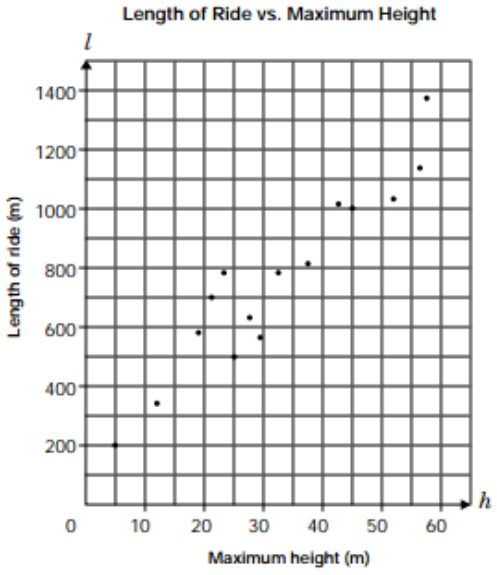
1. Thrill Rides

Susanna travels to different amusement parks to ride 15 roller coasters and collect data about each ride.

She constructs a scatter plot to show the relationship between the **total length** of the ride, l, in metres, and the **maximum height** of its peaks, h, in metres.

a) Draw a **line of best fit** to represent the data.

b) Determine an **equation** for your line of best fit. Justify your answer.



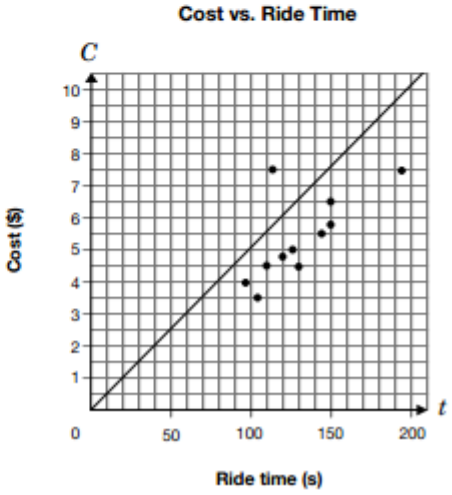
c) Susanna rides another roller coaster. The **length** of the ride on this roller coaster is **500 m.**

Determine its **maximum height,** using your results from part a) or b).
Justify your answer.

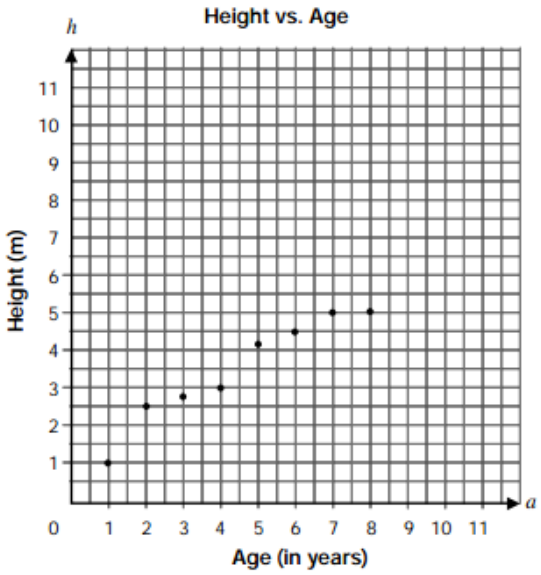
d) Susanna collects data about the relationship between the **cost of each ride,** C, in dollars, and **the time the ride lasts,** t, in seconds. She plots the data on the graph below.

Susanna graphs the equation $C = 0.05t$. She notices that its line is **not** the line of best fit.

Describe how to change the equation so that it represents the equation of a line of best fit for her data.
Justify your answer.



9. The graph below represents the relationship between the height, h, in metres, and the age, a, in years, of a tree.



What is the approximate **height** of the tree if it is **10 years old**?

a 10.5 m

b 8.5 m

c 6.5 m *

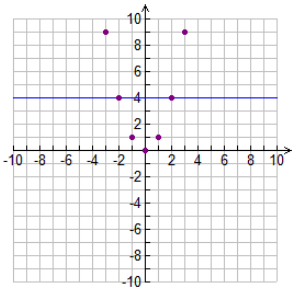
d 4.5 m

Review – Unit 5 – Statistics (Data Analysis)

Summary of the Main Ideas

The Process of Data Management

- Collect Data Involving Two Variables (e.g. measurements, surveys, etc.)
- Decide Which Variable is Independent and Which is Dependent
- Plot Data Points
  Create Scatter Plot → Don't connect the dots!!
  Independent Variable: Horizontal Axis
  Dependent Variable: Vertical Axis
- Look for Trends (e.g. upward, downward, no trend)
- Investigate how the Variables are Related
- Use Regression to Find "Curve of Best Fit"
  (Use a graphing calculator or programs such as TI-Interactive, Fathom, spreadsheets.)
- Since we are only familiar with linear relations, we are only concerned with "Lines of Best Fit" in this course.
- Linear models are only appropriate when the data follow a ***general upward trend*** or a ***general downward trend*** and the ***rate of change is roughly constant***. If this is not the case, some other model will be more appropriate.



Shown at the left is the line of best fit for the data points given in the scatter plot. Obviously, a linear model is completely inappropriate in this case.

Line of Best Fit: $y = 4$

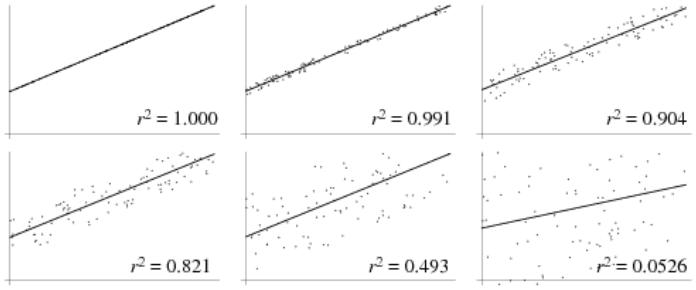
Correlation: $r = 0$

The Purpose of a Statistical Analysis

- The main objective of a statistical analysis is to determine how "strongly" two variables are related to each other.
- If two variables are strongly related, then regression is used to find an equation that best fits the data. The equation is then used to make ***predictions***.
- This process allows us to construct ***mathematical models*** of real-world situations.

How can we Tell if we have a Good Fit?

- The ***correlation coefficient*** (usually denoted r) is used to measure how "strongly" two variables are related.
- **Positive Correlation (Upward Trend)**
  The closer r is to 1, the better the fit.
  The closer r is to 0, the worse the fit.
- **Negative Correlation (Downward Trend)**
  The closer r is to –1, the better the fit.
  The closer r is to 0, the worse the fit.
- To avoid negative values, r^2 is often used instead of r.
  The closer r^2 is to 1, the better the fit.
  The closer r^2 is to 0, the worse the fit.
- The value of r can be calculated by using a graphing calculator or programs such as TI-Interactive and Fathom.



Important Skills

- Set scales on axes (use break if necessary)
- Distinguish between the independent and dependent variables.
- Plot data points to create scatter plot.
- Estimate the line of best fit using a ruler and pencil.
- Find the line of best fit using a graphing calculator or computer program.
- Identify ***outliers***. Outliers are measurements that differ significantly from the main body of the data. In other words, outliers are data points that do not follow the general trend of the bulk of the data.

Interpolation and Extrapolation

- **Interpolate**: Estimate a value ***between*** two measurements in a set of data. The ***average*** of the values is often used as an interpolated estimate.
- **Extrapolate**: Estimate a value ***beyond*** the range of a set of data.

Interpolated values are usually more reliable than extrapolated values. The reliability of extrapolated values tends to decrease as the distance from the data points increases.

Review Questions

True or False

1. _____ A scatter plot is a type of graph that can be used to find a relationship between two variables.

2. _____ Two values on a scatter plot can be used to extrapolate a value between them.

3. _____ If the ordered pairs on a scatter plot lie in a straight line, the relationship between the two variables is linear.

4. _____ A line of best fit must pass through all data points of a graph.

5. _____ The following set of ordered pairs has a linear relationship: (0, 0), (1, 1), (2, 4), (3, 9), (4, 16)

6. _____ This set of points has a non-linear relationship: (–6, –3), (–5, –1), (–4, 1), (–3, 3), (–2, 5), (–1, 7), (0, 9)

7. _____ Time is the dependent variable in a distance-time graph.

8. _____ In a distance-time graph, a horizontal straight line means that the speed is zero.

9. _____ A downward straight line in a distance-time graph of a car's movement means that the car is losing speed.

10. _____ A non-linear graph is produced in a distance-time graph when a person changes speed.

11. _____ The following ordered pairs have a non-linear relation: (0, 0), (1, 1), (2, 4), (3, 9), (4, 16), (5, 25)

Answers: T, F, T, F. F, F, F, T, F, T, T

Multiple Choice

1. A survey of favourite colours was conducted on 150 grade 9 students with the following results.

Colour	Red	Blue	Green	Pink	Yellow
Number	60	50	20	5	15

There are 500 students in grade 9. How many would you expect to choose blue as their favourite colour?

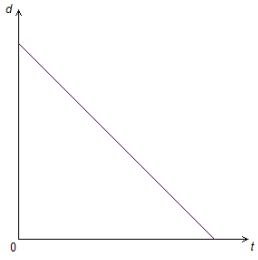
a. 50 b. 167 c. 100 d. 300

2. The table shows the height of a bean plant in the first week after it germinated. Predict the height of the bean plant on the eighth day.

Day	1	2	3	4	5	6	7
Height (cm)	0.9	1.8	2.6	3.6	4.4	5.5	6.1

a. 6.7 b. 7.5 c. 7.1 d. 8.0

3. The motion on this distance-time graph can be described as follows.



a. The person is not moving.
b. The person is walking toward the motion detector.
c. The person is walking away from the motion detector.
d. The person is slowing down.

4. Students in a grade 9 phys-ed class were surveyed about how often they exercise in a week.

Time	**Frequency**
Every day	10
Every other day	6
Only on weekends	14

If there are 1000 students in the school, how many do you expect exercise every day?

a. 268 b. 396 c. 333 d. 400

Answers: b, c, b, c

Problems

1. This table shows the numbers of days absent from mathematics class and the math marks for 15 students.

Number of Days Absent	Math Mark (%)
2	82
0	75
10	48
6	62
1	76
23	35
13	42
2	96
1	54
3	73
7	65
0	79
10	60
16	43
1	84

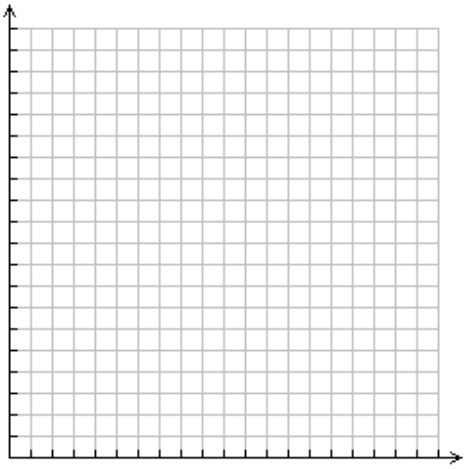
a) Identify the independent variable and the dependent variable. Explain your reasoning.

b) Make a scatter plot of the data.

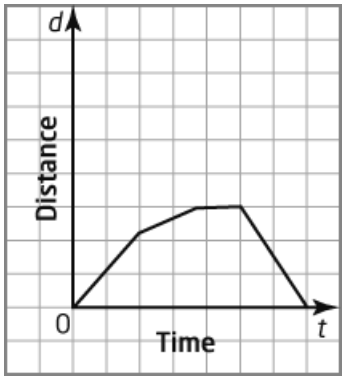
c) Describe the relationship between a student's marks and attendance.

d) Are there any outliers? If so, explain how they differ from the rest of the data.

d) Draw a line of best fit. Is a linear model appropriate? Explain.



2. Tom's movements after he left his house are shown on this distance-time graph. Describe his movements.



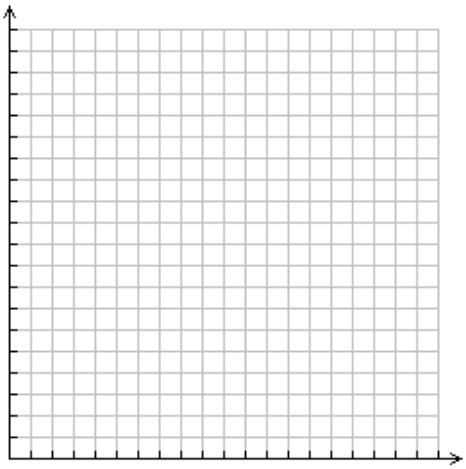
3. A skydiver jumps from an airplane. The distance fallen and time taken are recorded in the table.

Time (s)	Distance (m)
0	0
1	5
2	19
3	42
4	74
5	115

a) Draw a scatter plot of the relation.

b) Classify the relation as linear or non-linear. Explain your choice.

c) How far will the skydiver have fallen in 3.5 s?



4. The table shows the winning throws in the discus in the Olympic Games from 1956 to 2004.

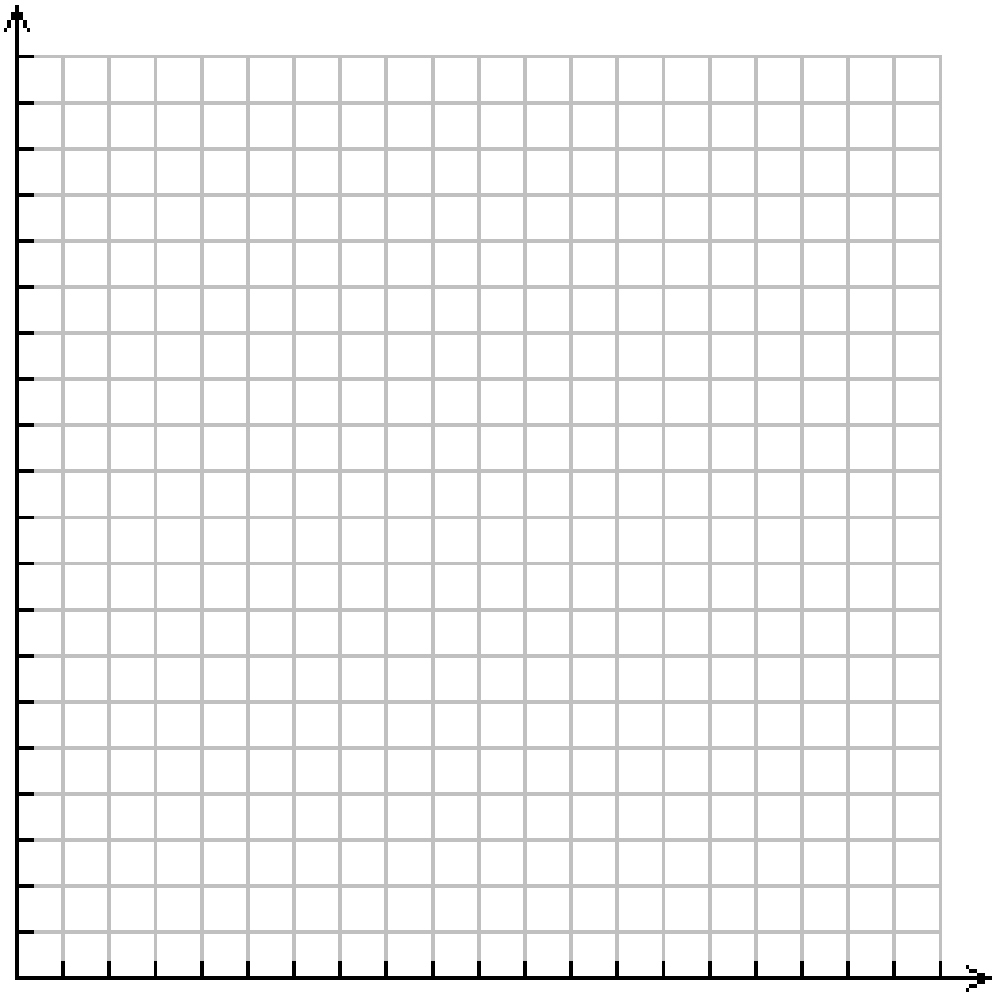
Year	Men's Distance (m)	Women's Distance (m)
1956	56.36	53.69
1960	59.18	55.10
1964	61.00	57.27
1968	64.78	58.28
1972	64.40	66.62
1976	67.50	69.00
1980	66.64	69.96
1984	66.60	65.36
1988	68.82	72.30
1992	65.12	70.06
1996	69.39	69.65
2000	69.30	68.40
2004	69.89	67.02

a) Graph both sets of data.

b) Use your graphs to compare any trends in the data. (Treat men separately from women.)

c) Identify any outliers. What may account for such outliers? Explain whether you should discount these outliers.

d) Use the data and the graphs to predict the winning distances for the men's and women's discus in the 2016 Olympic Games. Give reasons for your estimates.



5. This table gives the mean body temperatures of nine beetles at certain air temperatures.

Temperatures (°C)

Air	25.7	30.4	28.7	31.2	31.5	26.2	30.1	31.5	18.2
Body	27.0	31.5	28.9	31.0	31.5	25.6	28.4	31.7	18.7

a) Make a scatter plot of the data.

b) Identify the dependent and independent variables and determine what relationship, if any, exists in the data.

c) Draw a line of best fit for the data in the scatter plot. Is a linear model a good fit? Explain.

d) Predict the body temperature of a beetle at air temperature of
i) 27.5°C
ii) 35°C