MBF 3C Unit 5 – Statistics – Outline

Day	Lesson Title	Specific Expectations
1	One-variable data	D1.1 D1.2
2	Sampling Types and Techniques	D1.3 D1.4
3	Identify and Graphing One-Variable Data	D1.5
4	Common Distribution Properties and Questionnaire Design	D1.1 D1.6
5	Collecting and Organizing One-Variable Data	D1.1 D1.4
6	Measures of Central Tendency	D1.7 D1.8
7	Measures of Spread	D1.7 D1.8
8	Analyzing One-Variable Data	D1.9 D1.10
9	Review Day	
10	Test Day	
TOTAL DAYS:		10

D1.1 – identify situations involving one-variable data (i.e., data about the frequency of a given occurrence), and design questionnaires (e.g., for a store to determine which CDs to stock; for a radio station to choose which music to play) or experiments (e.g., counting, taking measurements) for gathering one-variable data, giving consideration to ethics, privacy, the need for honest responses, and possible sources of bias (Sample problem: One lane of a three-lane highway is being restricted to vehicles with at least two passengers to reduce traffic congestion. Design an experiment to collect one-variable data to decide whether traffic congestion is actually reduced.);

D1.2 – collect one-variable data from secondary sources (e.g., Internet databases), and organize and store the data using a variety of tools (e.g., spreadsheets, dynamic statistical software);

D1.3 – explain the distinction between the terms population and sample, describe the characteristics of a good sample, and explain why sampling is necessary (e.g., time, cost, or physical constraints) (Sample problem: Explain the terms sample and population by giving examples within your school and your community.);

D1.4 – describe and compare sampling techniques (e.g.,random,stratified,clustered,convenience, voluntary); collect one-variable data from primary sources, using appropriate sampling techniques in a variety of real-world situations; and organize and store the data;

D1.5 – identify different types of one-variable data (i.e., categorical, discrete, continuous), and represent the data, with and without technology, in appropriate graphical forms (e.g., histograms, bar graphs, circle graphs, pictographs);

D1.6 – identify and describe properties associated with common distributions of data (e.g., normal,bimodal,skewed);

D1.7 – calculate, using formulas and/or technology (e.g., dynamic statistical software, spreadsheet, graphing calculator), and interpret measures of central tendency (i.e., mean, median, mode) and measures of spread (i.e., range, standard deviation);

D1.8 – explain the appropriate use of measures of central tendency (i.e., mean, median, mode) and measures of spread (i.e., range, standard deviation) (Sample problem: Explain whether the mean or the median of your course marks would be the more appropriate representation of your achievement. Describe the additional information that the standard deviation of your course marks would provide.);

D1.9 – compare two or more sets of one-variable data, using measures of central tendency and measures of spread (Sample problem: Use measures of central tendency and measures of spread to compare data that show the lifetime of an economy light bulb with data that show the lifetime of a long-life light bulb.);

D1.10 – solve problems by interpreting and analysing one-variable data collected from secondary sources.

Unit 5 Day 1: Statistics - One Variable Data			MBF 3C
	Description Identify situations with one-variable data. Collect, organize and store data from secondary sources.		Materials Internet, Excel, Fathom, Stats Canada Handout or web-link BLM 5.1.1,5.1.2
		Assessment Opportunities	
Minds On…	**Pairs → Think /Pair/ Share** Ask students to think about what "Statistics" means to them. They then share with their partner, and finally with the class. Introduce the fact that all of these things we know about statistics will be explored in this unit. Post (or broadcast electronically) the annual average precipitation rates of Canadian and other international cities. Discuss possible uses for this information. BLM5.1.1 http://www40.statcan.ca/l01/cst01/phys08a.htm?sdi=precipitation (A hard copy is included.)		Real world applications might include farming, travel, tourism, real estate etc.
Action!	**Whole Class →Teacher Directed** **One-variable statistics lesson:** *Each column in the table from Statistics Canada represents a list of ***one-variable statistics***. This means that every ***entry*** (or number) in the column is measuring the same, single, unknown. *In tabular form, it can be difficult to identify trends in the data. To better understand your data, you need to sort and organize it. *This is done in two ways 1) Frequency Distribution Table 2) Histogram (Graph) **Frequency Distribution** By sorting data into intervals (or classes) and counting the number of entries that fall into each interval, it becomes easier to make a graph which allows us to quickly spot trends. ***Rules:*** 1. Too few or too many intervals will make it hard to analyze your data. Try to stick to 5-20 intervals. To do this, first find the range of data, and then divide that number by both 5 and 20 to determine how big each interval should be. 2. Make sure that the intervals don't overlap. If they do, you may end up counting some entries twice. To avoid this, add a decimal place to the start and end values of each interval. **Ex 1** a) Make a frequency distribution table to represent the number of wet days in Canadian cities by looking at the Stats Canada table. b) Make a histogram using your frequency distribution.		

Step 1: Find the range.

Range= Highest #- Lowest #

=217-109

=108

Interval Length:

5 intervals (bars) $= \frac{108}{5}$

$= 21.6$

20 intervals (bars) $= \frac{108}{20}$

$= 9$

∴ we want intervals anywhere from 9 units to 21.6 units wide.

To make counting easier, we choose any number between 9 and 21.6 that is easy to count by.

∴ good interval length = 20 (this could be any other number such as 10 or 15)

Step 2: Avoid overlap. Add a decimal to the start and end values of each interval.

To choose a starting interval, be sure that it includes the lowest number (in this case 109).

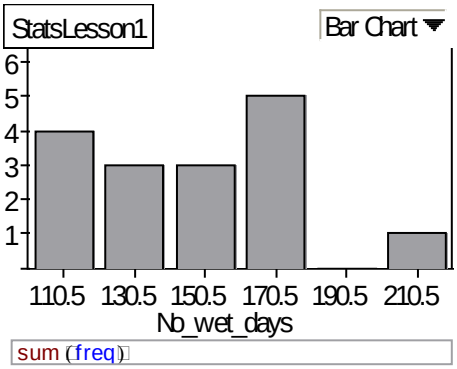
∴ good starting interval is 100.5-120.5 (note: this is 20 units long with an extra decimal place added)

Step 3: Sort the data in a table

Interval	Tally	Frequency	Cumulative Frequency
100.5-120.5	IIII	4	4
120.5-140.5	III	3	7
140.5-160.5	III	3	10
160.5-180.5	IIIII	5	15
180.5-200.5		0	15
200.5-220.5		I	16

Note: *Keep counting your intervals by twenty until you've included the last number (in this case, 217).
*A cumulative frequency column in a good way to double check that you didn't miss any entries.

b) **Organize data in graphical form.**



Notes: *The y-axis is frequency

If you are planning on graphing in Fathom, note:
1. As you are dropping the x column onto the empty graph, hold down the shift key just before you let it drop, this will create a bar graph.
2. You must change the formula at the bottom from "count" which is the default to "sum" by clicking on the word count and typing over it.

A break should be included in

		*The x-axis represents whatever you are counting *Unless your interval starts at zero, you should include a break in your graph *It is often easier to write the midpoint of each interval rather than the start and end points *There are no spaces between the bars since the intervals are **continuous**, this means that there is no break in the x-values <u>Demonstrate how to import data from the internet or Excel to Fathom (or both)</u> ***1. From the internet to Fathom*** a) Open a new document in Fathom, click on "File", then on "Import from URL" and type in the address of the website that you want. Or b) Open a new document in Fathom, also open the website that you want so that both windows appear on-screen at once. Click on the web address and drag it into the Fathom document. ***2. From Excel to Fathom*** a) In Excel, use the mouse to select all of the cells that you want. While selected, copy them (Ctrl-C) or right click and copy. b) Open Fathom, drop a new collection box into it and click "Edit", then on "Paste Cases".		the x-axis.
	Consolidate Debrief	**<u>Whole Class → Discussion</u>** Ask the students to summarize what they now know about statistics.		
Application		**Home Activity or Further Classroom Consolidation** Students complete BLM 5.1.1		

MBF3C
BLM5.1.1

Statistics Canada-Precipitation Data

Related tables: Weather conditions.

Weather conditions in capital and major cities

(Precipitation)

	Annual average		
	Snowfall	**Total precipitation**	**Wet days**
	cm	mm	number
St. John's	322.1	1,482	217
Charlottetown	338.7	1,201	177
Halifax	261.4	1,474	170
Fredericton	294.5	1,131	156
Québec	337.0	1,208	178
Montréal	214.2	940	162
Ottawa	221.5	911	159
Toronto	135.0	819	139
Winnipeg	114.8	504	119
Regina	107.4	364	109
Edmonton	129.6	461	123
Calgary	135.4	399	111
Vancouver	54.9	1,167	164
Victoria	46.9	858	153
Whitehorse	145.2	269	122
Yellowknife	143.9	267	118
International comparisons			
Beijing, China	30	623	66
Cairo, Egypt	...	22	5

Capetown, South Africa	...	652	95
London, England	...	594	107
Los Angeles, U.S.A.	...	373	39
Mexico City, Mexico	...	726	133
Moscow, Russia	161	575	181
New Delhi, India	...	715	47
Paris, France	...	585	164
Rio de Janeiro, Brazil	...	1,093	131
Rome, Italy	...	749	76
Sydney, Australia	...	1,205	152
Tokyo, Japan	20	1,563	104
Washington, D.C.	42	991	112

... : not applicable.
Sources: For Canada, *Climate Normals 1961–1990*, Climate Information Branch, Canadian Meteorological Centre, Environment Canada; for International data, *Climate Normals 1951–1980.*
Last modified: 2005-02-16.

MBF3C Name:
BLM 5.1.2 Date:

Statistics Work

1. Create a frequency distribution and histogram for each of the following using the data from Stats Canada:

a) Annual average precipitation (mm) in Canada
b) Annual average precipitation (mm) in international cities
c) Number of wet days in international cities

2. a) Go to the World Cup of Soccer website (www.world-cup-info.com/statistics/world_cup_games_played.htm) and enter the data into Fathom.
b) Create a graph of the number of points scored per country by dragging the graph icon into your document and dragging the needed columns from your case table.

Or

a) Go to the Toronto Maple Leafs website (http://mapleleafs.nhl.com/club/roster.htm) and enter the data into Fathom.
b) Create a graph with the x-attribute representing the number of games played (GM) and the y-attribute representing the points per game (PPG)

Write a concluding statement based on your graph.

Questions 3 to 7 are based on the following information.

The pulses of 30 people were taken for 1 minute and recorded. These are the results:

66	79	53	81	84	76	76	67	64	83
92	56	67	77	91	61	71	86	73	87
71	67	71	81	86	62	77	91	72	68

3. Why is it hard to spot the trends in the data as it appears?

4. a) Make a frequency distribution table for the above data including a cumulative frequency column. Start with 50.5-55.5 as your first interval.
b) Construct a histogram based on your frequency distribution.

5. Use your graph to answer each question:
a) In which interval does the most common pulse occur?
b) In which interval does the least common pulse occur?

6. What percentage of the people have a pulse over 85.5?

MBF3C Name:
BLM 5.1.2 Date:

Statistics Work (continued)

7. a) If you record the pulse for 300 people, how many would you expect to have a pulse in the interval 75.5-80.5? Give reasons for your answer.
b) What assumptions are you making?

Questions 8 to 12 are based on the following information.

An English class had the following grades on a test (out of 100).

26	63	73	82	32	73	35	63	56	87
40	51	55	43	53	70	43	92	64	75
46	64	23	67	52	28	76	56	67	

8. Start with the interval 20.5-30.5. Create a frequency distribution.

9. a) Create a histogram.
b) Which interval has the greatest frequency?

10. a) What percentage of the class received an A (80% or better)?
b) What percentage of the class failed (under 50%)?

11. The same class wrote a second test. These are their marks.

66	62	14	41	45	89	59	43	67	37
31	65	50	43	53	57	54	84	68	74
61	54	34	70	45	64	76	70	65	

Repeat questions 8 and 9 for this set of marks.

12. Compare the two histograms created.
a) What differences are there?
b) What similarities are there?
c) What information do the differences indicate to the teacher?

MBF3C

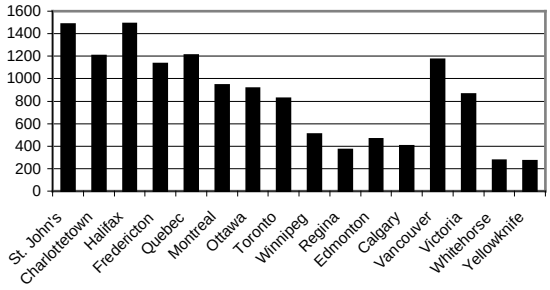
BLM 5.1.1

Statistics Work Solutions

1. a) i)

Annual average precipitation (mm)	Frequency (f)
St. John's	1482
Charlottetown	1201
Halifax	1474
Fredericton	1131
Quebec	1208
Montreal	940
Ottawa	911
Toronto	819
Winnipeg	504
Regina	364
Edmonton	461
Calgary	399
Vancouver	1167
Victoria	858
Whitehorse	269
Yellowknife	267

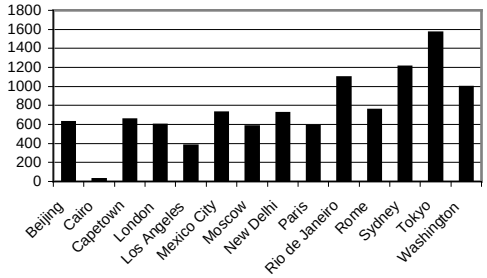
ii)



b) i)

Annual average precipitation (mm)	*Frequency* (f)
Beijing	623
Cairo	22
Capetown	652
London	594
Los Angeles	373
Mexico City	726
Moscow	575
New Delhi	715
Paris	585
Rio de Janeiro	1093
Rome	749
Sydney	1205
Tokyo	1563
Washington	991

ii)



MBF3C

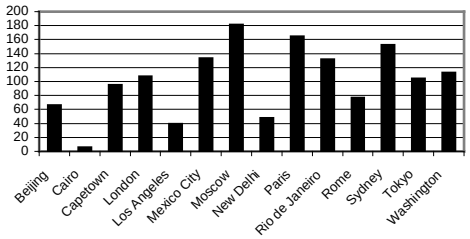
BLM 5.1.1

Statistics Work Solutions

c) i)

Number of wet days (x)	Freq. (f)	Number of wet days (x)	Freq. (f)
Beijing	66	New Delhi	47
Cairo	5	Paris	164
Capetown	95	Rio de Janeiro	131
London	107	Rome	76
Los Angeles	39	Sydney	152
Mexico City	133	Tokyo	104
Moscow	181	Washington	112

ii)

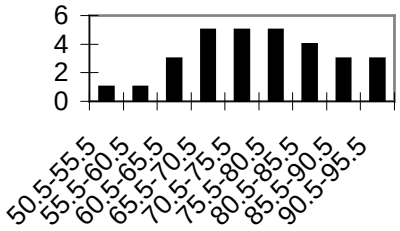


3. Not organized or ranked; difficult to compare data.

4. a)

Number of pulses (x)	Frequency (f)	Cumulative frequency
50.5-55.5	1	1
55.5-60.5	1	2
60.5-65.5	3	5
65.5-70.5	5	10
70.5-75.5	5	15
75.5-80.5	5	20
80.5-85.5	4	24
85.5-90.5	3	27
90.5-95.5	3	30

b)



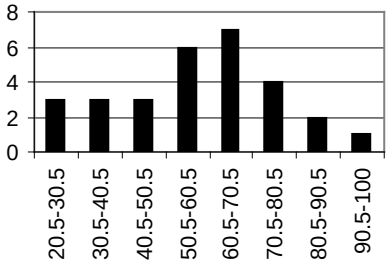
5. a) 65.5-70.5; 70.5-75.5; 75.5-80.5 **b)** 50.5-55.5; 55.5-60.5 **6.** 20%

7. a) 50; Determine the percent frequency of the interval and multiply by the total number of people. **b)** Answers may vary; for example: pulses do not change.

8.

Grades (x)	Frequency (f)
20.5-30.5	3
30.5-40.5	3
40.5-50.5	3
50.5-60.5	6
60.5-70.5	7
70.5-80.5	4
80.5-90.5	2
90.5-100	1

9. a)



b) 60.5-70.5

MBF3C

BLM5.1.2

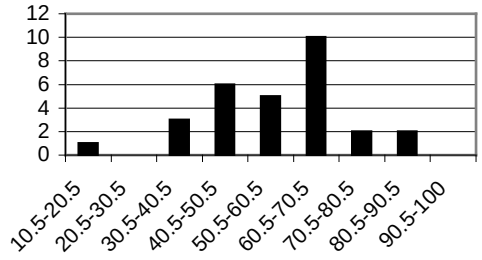
Statistics Work Solutions

10. a) 10.3% **b)** 31.0%

11. i)

Grades (x)	Frequency (f)
10.5-20.5	1
20.5-30.5	0
30.5-40.5	3
40.5-50.5	6
50.5-60.5	5
60.5-70.5	10
70.5-80.5	2
80.5-90.5	2
90.5-100	0

ii)



iii) 60.5-70.5

12. a) Answers may vary; for example: second test resulted in lowest grade (14).
b) Answers may vary; for example: failure rate. **c)** Answers may vary.

Unit 5 Day 2: Statistics - Sampling		MBF 3C
	Description Sampling Types and Techniques Explain the distinction between population and sample, providing relevant examples. Describe and compare sampling techniques	Materials BLM 5.2.1
		Assessment Opportunities
Minds On…	**Whole Class →Discussion** Pose the following statement to the students. Nathalie Beauchamp surveys randomly from her on-line youth book club members as well as the lists of youth cardholders at the two nearest community libraries. She returns to school and suggests to her friend on students' council that the school should host a read-a-thon to raise money for prom since the participants in her survey all felt that it was a good idea. What is the problem with her research?	Possible discussions could include how Nathalie only surveyed people who would be more likely to participate since they are active readers already.
Action!	**Whole Class →Teacher Directed** **Sampling Types and Techniques Lesson:** **Note:** Nathalie surveyed only some people and used their feedback to make a general statement about a larger group (i.e. all students at her high school). In this example, the ***population*** is high school students since that is the group about which she made the statement. The ***sample*** is the group of people that she chose to survey. This includes the book club and library respondents. In general, the ***population*** is the entire group being studied and the ***sample*** is the group of people taken from that population. **Advantages and Disadvantages**: A population, if surveyed, will give you really accurate results, but it is often very hard to ask everybody in a population (i.e. all high school students). * If everyone in a population *is* surveyed, then it's called a ***census***. A sample is easier to find and survey, but your results may be biased. This means that you could be misled based on who you surveyed if the group didn't accurately represent the population. **Sampling Techniques:** **Random Sample** *In a simple, random sample, all selections are equally likely. **E.g.:** Drawing 5 names from a hat holding 30 names and surveying those 5 people. **Pros:** Easy to do. Fair to all involved. **Cons:** Could get a poor representation of the population. i.e. All 5 names drawn could be close friends who share the same opinion on everything.	

Stratified Sample
*The population is divided into groups, then a random sample is taken of each group.
*The number sampled from each group is proportional to the size of the group.

E.g.: A school is divided into 4 groups by grade. There are 300 grade nines, 350 grade tens, 270 grade elevens and 320 grade twelves. Proportion of each group chosen → 10%

Thirty grade nines are surveyed, 35 grade tens, 27 grade elevens and 32 grade twelves.

Pros: A fair representation of the population.

Cons: Takes more work to set up, can still be biased.
i.e. If the survey is about driving permits, the grade eleven and twelve students may respond differently.

Cluster Sample
*The population is divided into groups.
*A random number of groups is chosen. (It could be just one group).
*All members of the chosen group(s) are surveyed.

E.g.: A VP enters the cafeteria and randomly selects two tables. All students at those two tables are surveyed.

Pros: Easy to do.

Cons: Often over-represent some opinions and under-represent others.

Convenience Sample
*A selection from the population is taken based on availability and/or accessibility.

E.g.: To survey woodworkers in Ontario, we ask people at several lumber yards and home improvement stores scattered about the province.

Pros: A good way to gain ideas when you're starting to research an idea.

Cons: You have no idea how representative your sample is of the population.

Voluntary Sampling
*People volunteer to take part in a study.

E.g.: Psych 101 students at Trent University are given an additional 2% at the end of the year if they volunteer for any two upper-year psychology surveys and/or studies.

Voting on Canadian Idol.

Pros: Often useful for psychological and/or pharmaceutical trials.

		Cons: Sometimes (as in TV voting), participants can vote more than once and/or be surveyed more than once, skewing the results.		
	Consolidate Debrief	**Pairs → Think/ Pair/ Share** The class can orally give an example of each type of study that they've either participated in or are familiar with due to the media.		
Application Concept Practice		**Home Activity or Further Classroom Consolidation** Students complete BLM 5.2.1		

MBF3C Name:

BLM 5.2.1 **Sampling** Date:

1. In order to find out which songs are the most popular downloads, a survey was sent out to a number of teenagers.
    a) What are some advantages of using a survey to collect data?
    b) What are some disadvantages to this method?
    c) What would be another way to get this same information?

2. Sometimes it is better to ask all of the population before making a decision. For each scenario, state whether a **sample** should be used or a **census**.
    a) Testing the quality of the air in airplanes.
    b) Determining the popularity of a particular website.
    c) Determining the number of potential buyers of a new MP3 player.
    d) Determining the chemical composition of a good barbeque sauce.
    e) Checking the air pressure of the tires on a car.
    f) Determining the effectiveness of a new laser-eye surgery.

3. Given the following four options, which would be most effective in predicting the outcome of the upcoming municipal election for mayor, and why?
    a) 100 completed surveys that were handed out randomly through the city.
    b) 100 phone calls made to different parts of the city.
    c) 100 people interviewed at a local neighbourhood-watch party.
    d) 100 surveys completed by children at a local middle school.

4. A school board received a load of 10 000 graphing calculators to pass out to their high schools. They were concerned with the state of the delivery and therefore with the number of defective calculators. They decided to check them out.
    First, 20 calculators were checked and all worked perfectly.
    Second, 100 calculators were tested and 2 were broken.
    Third, 1000 were tested and 15 were broken.

    a) After the first test, would it be fair to say that none of the calculators were broken? Why or why not?
    b) Whose statement is likely more accurate?
        Sami: 2% are defective Sima: 1.5% are defective
    c) In the shipment of 10 000, how many would you estimate to be defective? Explain.

5. Gelman's Rent-All want to see if they should open up a second shop at a neighbouring plaza. They conduct a poll by leaving sheets at the entrance of the plaza and asking people to fill them in.

    a) What type of sample is this?
    b) What are some of the pros of this method?

MBF3C Name:

BLM 5.2.1 **Sampling (continued)** Date

6. A local high school has 600 students in grade 9, 400 in grade 10, 300 in grade 11 and 200 in grade 12. A sample of 100 students is used to choose which brand of chocolate bar should be sold in the vending machine. How many of the 100 surveys should be handed out to
   a) Grade 10s?
   b) Grade 11s?
   c) What type of sample is this?

7. For each scenario, state whether a stratified sample should be used. Explain your reasoning.
   a) Canada wants to hold a general referendum to decide a major political issue. A sample of 10 000 people is chosen to predict the outcome.
   b) A shipment of 35 000 clear plastic rulers is to be checked for defects.
   c) There are 250 women and 750 men working at Harpo studios. A sample of 20 is taken to determine what type of end-of year party should be planned.
   d) The director of a local community centre is supposed to decide if any of her budget should be spent on pool maintenance.
   e) At a Tai Chi club, an opinion poll is to be conducted on the quality of the equipment.

8. For each of the following samples, the cluster technique was used. Which would result in a fair sample (F) and which would result in a poor sample (P)?
   a) Asking ER-nurses about the value of a new triage approach.
   b) Going to a high school to determine the most popular brand of jeans.
   c) Asking only senior students about the prom location.
   d) Asking Smart-Car owners about a hot environmental issue.

Solutions

1. a) answers may vary; for example: easy to conduct. **b)** answers may vary; for example: may not be representative of entire population. **c)** answers may vary; for example: interview, case study. **2. a)** sample **b)** census **c)** sample **d)** sample **e)** census (not all tires have the same pressure necessarily) **f)** sample
3. a) random; variety in responses will reflect different viewpoints
4. a) no; for example: 20 is not a representative value of 10 000. **b)** 1.5%
c) 150; based on 3rd method which is more accurate. **5. a)** convenience
b) answers may vary; for example: gather helpful ideas; is not time consuming to conduct. **6. a)** approx. 27 **b)** 20 **c)** stratified **7.** answers may vary **a)** yes; samples will be proportional to the total constituents **b)** no; ineffective **c)** yes; represents both sexes fairly **d)** no; director should consult board members and those directly related **e)** no; use other sample technique **8. a)** P **b)** F **c)** P **d)** F

Unit 5 Day 3 :Statistics - Graphing		MBF 3C
	Description Identifying and Graphing One-Variable Data *Identify discrete, continuous and categorical data and represent in graphical form with and without technology	Materials Graphing Calculator, Protractor, Fathom or Excel BLM 5.3.1

			Assessment Opportunities
Minds On…	**Whole Class → Discussion** Look at these 3 questions. How is the data collected by each question different? 1. Please check the reason for your previous work absence. X Illness X Vacation and/or Holiday X Funeral 2. How many km/L of gas does your current car get per tank? 3. How many years of schooling does your career require? 1=Categorical, 2=Continuous, 3=Discrete		Here is an opportunity to discuss an understanding of qualitative vs. quantitative data and when each might be better.
Action!	**Whole Class → Teacher Directed Lesson** **Identifying and Graphing One-Variable Data Lesson:** *Data can be recorded in several different ways; there are three types that we are going to look at. **1. Categorical Data** (Qualitative) *This is data which is usually recorded as a label and not a number. e.g. #1 in Minds On **E.g.:** i) Checking male/female on a survey ii) Listing the type of car that you drive iii) Eye colour *Sometimes, categorical data *is* recorded as a number, but the value of the number is not as important as what it represents. *A common example of this is known as the ***Likert Scale***. This is frequently used on surveys where: 1=Strongly disagree 2=Disagree 3=Neutral 4=Agree 5=Strongly agree **2. Continuous Data** *This is numerical (or quantitative) data where values can exist between recorded values. i.e. decimals are allowed **E.g.:** Any measurement (mL, cm, m, weight, time, temperature) where decimals are permitted.		

3. Discrete Data
*This is also numerical data, but decimals are not allowed. There is a fixed number of possible values.

E.g.: Number of toppings on a pizza, money in cents, hockey scores

Ex 1:
For each, state the data type.

a) Number of mugs of coffee drank in a day.
b) Type of pet at home (e.g. dog, cat, bird, rodent, reptile)
c) Number of pets at home.
d) Amount of coffee in mL drank in a day.

Answer:
a) Discrete
b) Categorical
c) Discrete
d) Continuous

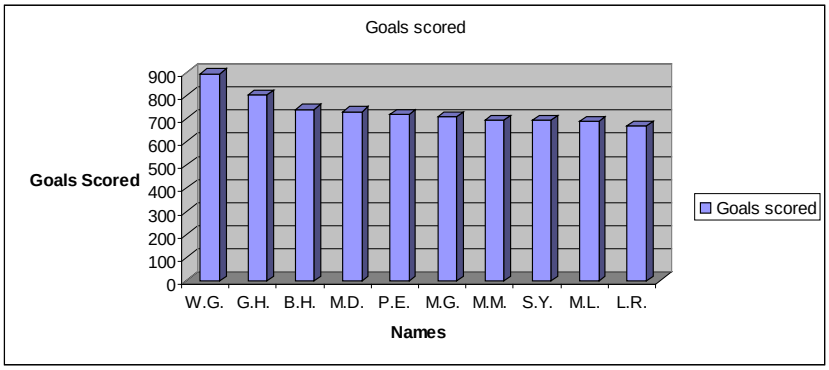
*There are several different types of graphs used to represent each data type.

1. Histogram (as previously seen in this unit)
*****Bars are used to represent continuous data. They touch since there are no breaks in the data.

2. Bar Graph
*Similar to a histogram except there are spaces between the bars.
*Used for discrete data.

Ex 2: Create a bar graph of the following hockey all-time regular season goal scorers.

Name	Goals scored
Wayne Gretzky	894
Gordie Howe	801
Brett Hull	741
Marcel Dionne	731
Phil Esposito	717
Mike Gartner	708
Mark Messier	694
Steve Yzerman	692
Mario Lemieux	690
Luc Robitaille	668



3. Circle Graph

*Used for numerical data when examining data in proportion to a whole.
*Good to see budgets, mark breakdowns, costs in manufacturing etc.

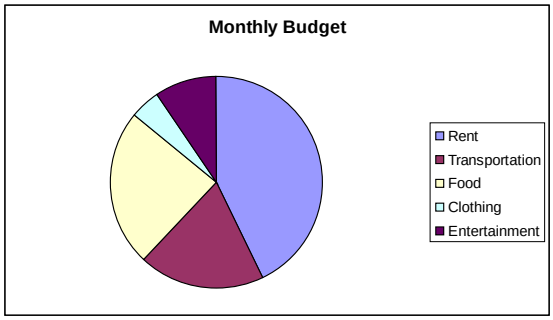
Ex 3: Given the following monthly budget, create a circle graph.

Item	Cost ($)
Rent	900
Transportation	400
Food	500
Clothing	100
Entertainment	200

Step 1: Add an additional row for each column total.
Add an additional column to calculate the corresponding degrees in a circle graph.

Item	Cost ($)	Degrees in Circle
Rent	900	$\frac{900}{2100} \times 360 = 154.3°$
Transportation	400	$\frac{400}{2100} \times 360 = 68.6°$
Food	500	$\frac{500}{2100} \times 360 = 85.7°$
Clothing	100	$\frac{100}{2100} \times 360 = 17.1°$
Entertainment	200	$\frac{200}{2100} \times 360 = 34.3°$
Total	2100	360°

Step 2: Mark a centre to your circle and draw a starting line.
From your starting line, use a protractor to measure each angle. Draw a line at each measurement and label each section.



4. Pictograph

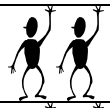
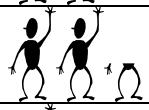
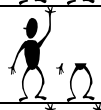
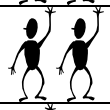
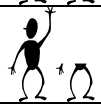
*Good way to visualize data.
*Not as precise as the other graphing styles.
*Commonly used for frequency with discrete data.

Ex 4: Draw a pictograph to represent the following list of students in clubs using the legend that 1 stick man= 25 people.

Football= 50 students
Band= 63 students
Soccer= 37 students
Musical Theatre= 52 students
Track= 35 students

Step 1: Create a table where the left column lists the possible activities and the right column will have the proper number of pictures. This is done as an estimate only.

Step 2: Draw the number of stick figures that most closely relates to the number of students in each activity.
Notice that 12 or 13 people can be represented by half of a stick figure.

Activity	No. of Participants
Football	
Band	
Soccer	
Musical Theatre	
Track	

Repeating some exercises with the use of technology:

Ex 5: As a class, count the number of students with each eye colour:

1. Brown
2. Blue
3. Green
4. Hazel

Create a bar graph using the graphing calculator.

Step 1: Enter the information into the calculator.

Go to ***Stat 1:Edit***.
In L_1, enter the numbers 1-4 representing each eye colour.
In L_2, enter the number of people for each.

Step 2: Graph it.

Turn on Stat Plot (go to ***2nd Y=***). Press ***Enter*** to turn on Plot 1.

Choose the graph type with your cursor. We want the bar graph icon.

Keep going until the bar graph is selected.
Press ***Graph***.
If the window needs adjustment, ***Zoom 9:Stat*** will always automatically adjust to any stat plot entries.

TI-89 Titanium Instructions

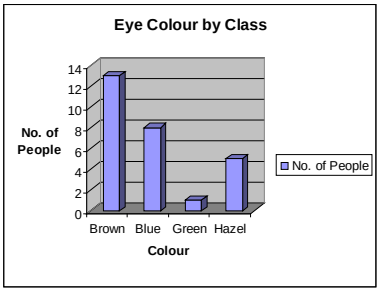
1. Go to Stats/List Editor Program from the APPS list
2. Enter the data into List 1 → 1 – 4 for the eye colours and List 2 → frequencies for each eye colour as surveyed.
3. Press F2 button → Plots
4. Select 1 → Plot Setup
5. Press F1 button while selecting Plot 1 (or any free plot)
6. In the Plot Type area select 4: Histogram
7. For the x data you need to move the cursor into that box and press 2nd button and then the minus button (VAR-LINK) then scroll down until you find "list 1" which you then select and press enter
8. On the "Use Freq and Categories?" line select Yes
9. Now in the "Freq" box you need to move the cursor into that box and press 2nd button and then the minus button (VAR-LINK) then scroll down until you find "list 2" which you then select and press enter
10. Press enter to accept the settings
11. Now press F5 button for ZoomData

Ex 6: Repeat the previous example using Fathom or Excel.

Excel

Step 1: Type the eye colours in column A.
Type the number of people in column B.
Click on the chart icon.
Select the style that you want. Keep clicking on Next until finished.

Colour	No. of People
Brown	13
Blue	8
Green	1
Hazel	5



Consolidate Debrief

Pairs → Think /Pair /Share

Have students think about what they now know about discrete, continuous and categorical data, share it with a partner, and then the teacher can ask for examples and help to clarify when needed.

Application
Concept Practice

Home Activity or Further Classroom Consolidation

Students complete BLM 5.3.1

MBF3C Name:
BLM 5.3.1 Date:

Graphing

1. The size of your school is stated in several different ways. For each measurement, state if it is discrete, continuous or categorical.
   a) The height of the building
   b) The number of rooms
   c) The number of floors
   d) The sum of the areas on each floor in m^2

2. Identify each variable as discrete, continuous or categorical.
   a) favourite TV show
   b) paint colour in bedroom
   c) English grade
   d) volume of IPod
   e) age
   f) calories in a meal
   g) number of cancer deaths last year
   h) monthly unemployment rate

3. The following is a list of ways to state the size of a book. Which variable can be continuous?
   a) thickness of binding used
   b) number of words
   c) length of pages
   d) number of pages

4. An emergency room technician assesses each patient that comes in. She records the following for each: blood pressure, age, gender, number of previous ER visits in the current year.
   a) How many of these variables are likely measured as continuous variables?
   b) How many are discrete?

5. Paul is determining whether or not his privately owned gas station will make it in his town. He asks automobile owners to name the station where they last bought gas. Is the variable that he is measuring discrete, continuous or categorical?

6. a) Construct a circle graph to illustrate the following data. First, you may want to complete the table.
   b) Construct a bar graph with the same data.

Percentage of Canadian Travellers Passing Through CanFly Airport per Day by Age

Age (Years)	Percent	Degrees in Circle
0-12	8	
13-18	12	
18-25	17	
26-40	32	
41-65	24	
65+	7	
Total		

7. Construct a pictograph to represent the following data.

Percentage of People Who Believe in Each Creature

Creature	People (%)
Loch Ness Monster	32
Big Foot	44
Ogopogo	25
Aliens	78

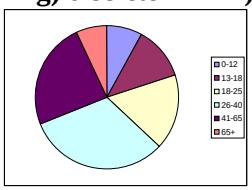
MBF3C
BLM 5.3.1

Name:
Date:

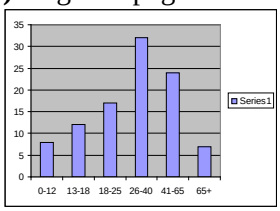
Graphing Solutions

1. a) continuous **b)** discrete **c)** discrete **d)** continuous **2.a)** categorical **b)** categorical
c) discrete **d)** continuous **e)** discrete **f)** discrete **g)** discrete **h)** continuous

3.c) length of pages **4.a)** 0 **b)** 1 **5.** categorical **6. a)**



b)



7. Answers will vary

Unit 5 Day 4: Statistics - Types of Distributions

MBF 3C

Description

Common Distribution Properties and Questionnaire Design
Identify and describe properties with common distributions.
Identify normal, bimodal and skewed distributions

Materials
BLM 5.4.1

Assessment Opportunities

Minds On…

Whole Class → Three Corners Activity

Each corner of the room is labelled (circle/pie chart, Bar graph, line graph)

For each question, determine the type of graph that would work best, move to that corner of the room and be prepared to defend your answer as a group.

a) What portion of students buy lunch every day? Fridays only?

b) What type of burger do people prefer: beef, chicken, turkey, or veggie?

c) How many calories are burned doing each activity for an hour: running, swimming, volleyball, golf?

Answers: a) Circle/ pie chart b) Circle/ pie chart c) Bar graph

Action!

Whole Class → Teacher Led Lesson

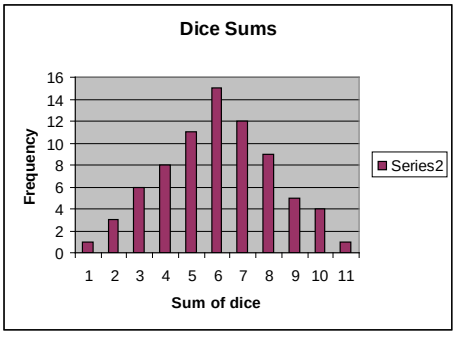
Common Distribution Properties Lesson:
Histograms can take on any of several common shapes. Among these distributions are both ***symmetrical*** and ***skewed*** graphs.

Part 1: Symmetrical Distributions
*Symmetrical distributions can be either ***normal***, ***bimodal*** or ***uniform***.

1. Normal Distributions
*These are commonly referred to as bell-curves or mound-shaped distributions.
*The middle interval(s) will have the greatest frequency (i.e. the tallest bar).
*All other intervals will have decreasing frequencies as you move away from the centre of the graph (i.e. the bars get smaller as you move out to the edges).

Ex 1: A pair of dice were rolled 75 times. After each roll, their sum was recorded and graphed.

Sum on dice	Frequency
2	1
3	3
4	6
5	8
6	11
7	15
8	12
9	9
10	5
11	4
12	1



It may be necessary to review the definition of symmetry here.

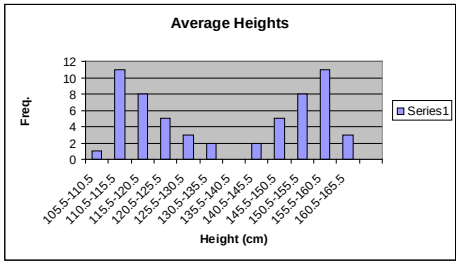
Note: Even though it isn't perfectly symmetrical, it still fits the definition of a normal distribution.

2. Bimodal Distributions

*These look like inverted normal distributions.
*The intervals with the highest frequencies (i.e. tallest bars) are at either end of the graph and the interval with the lowest frequency is in the centre.
*Frequencies increase as you move away from the centre of the graph.

Ex 2: A class of grade 6 and grade 1 students each measured their heights. They recorded and graphed them.

Height (cm)	Freq.
105.5-110.5	1
110.5-115.5	11
115.5-120.5	8
120.5-125.5	5
125.5-130.5	3
130.5-135.5	2
135.5-140.5	0
140.5-145.5	2
145.5-150.5	5
150.5-155.5	8
155.5-160.5	11
160.5-165.5	3

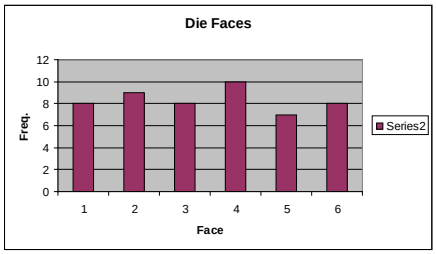


3. Uniform Distributions

*The frequencies of each interval are approximately equal.

Ex 3: A die is rolled 50 times. The face is recorded and graphed.

Die Face	Freq.
1	8
2	9
3	8
4	10
5	7
6	8



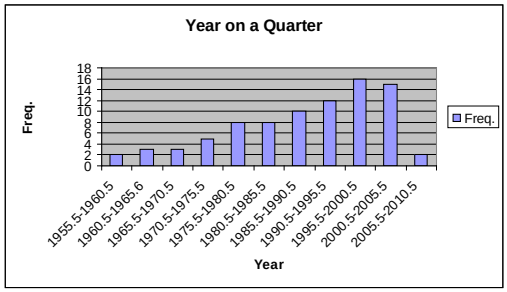
Part 2: Skewed Distributions

*There are 2 kinds of skewed graphs:

1. In ***right-skewed*** graphs, the bars with the highest frequencies are on the left side and the frequencies decrease as you move right.
2. In ***left-skewed*** graphs, the bars with the highest frequencies are on the right side and the frequencies decrease as you move left.

Ex 4: Sally picked up a handful of quarters. She recorded the year of each and made a graph.

Year	Freq.
1955.5-1960.5	2
1960.5-1965.6	3
1965.5-1970.5	3
1970.5-1975.5	5
1975.5-1980.5	8
1980.5-1985.5	8
1985.5-1990.5	10
1990.5-1995.5	12
1995.5-2000.5	16
2000.5-2005.5	15
2005.5-2010.5	2



Note: Even though there is a low-frequency bar on the right side, the trend is still ***left-skewed***.

Questionnaire and Experiment Tips:
Great questionnaires and experiments follow a few simple rules:

1. Give consideration to privacy. i.e. try to avoid asking for any personal information which is irrelevant for your survey.

2. Do not lead respondents' answers in order to prove a point. i.e. avoid asking a question such as *"Statistics show that avoiding white sugars and flours will greatly improve one's overall health. In how many daily meals, do you think that these ingredients should be included?"*

3. Experiments need to be done so that the question at hand is best addressed. i.e. The time of day must be applicable. Reviewing after-school activities should be done from 3:00 until 6:00.

Discussion options: When is it okay to ask someone's age? Weight? Income?

Consolidate Debrief

Whole Class → Discussion

When you do a survey what types of things should you think about in terms of how the survey is made up and what types of things you should be aware of in the graphical representations of the answers.

Application
Concept Practice

Home Activity or Further Classroom Consolidation

Students complete BLM 5.4.1

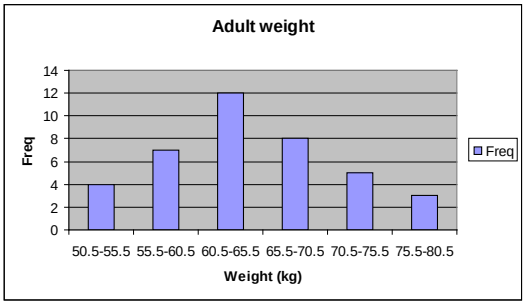
Questions 2 and 3 may be done as assignment questions. A questionnaire follow-up will occur next day.

MBF3C Name:

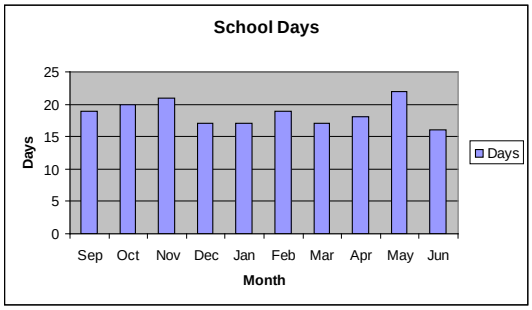
BLM 5.4.1 **Types of Distributions** Date:

1. Label each graph as ***normal, bimodal, uniform, left-skewed*** or ***right-skewed***.

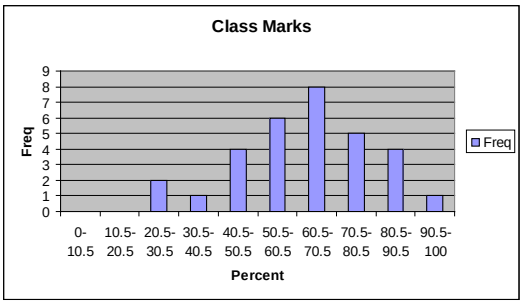
a)



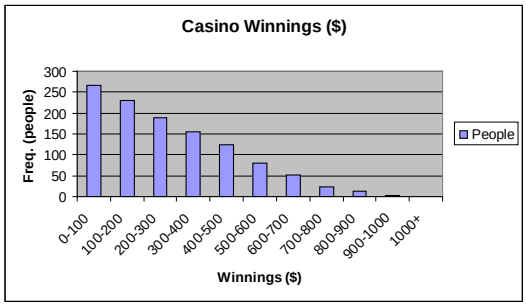
b)



c)



d)



2. Create a 5 question survey on one of the following topics:
   a) professional athlete earnings
   b) top movie songs
   c) favourite fair ride
   d) favourite summertime activity

3. The left-hand lane of the 401 express corridor is being restricted to vehicles with at least two passengers in order to reduce traffic congestion. **Outline your strategy for collecting data to determine if the traffic congestion is really reduced.**
   i) Include at least 3 steps in your process.
   ii) Identify the time(s) that you would conduct this experiment.
   iii) Indicate how you would record your data.
   iv) Indicate how you would use your data to answer the question.

MBF3C :
BLM 5.4.1

Types of Distributions Solutions

1. a) normal **b)** uniform **c)** right-skewed (discuss why it may appear to be normal) **d)** left-skewed **2.** answers may vary; for example: **c)** immigration: **i)** From which country have you originated? **ii)** What is the main reason for which immigrants leave their place of origin? **iii)** What is the main reason for which immigrants come to Canada? **iv)** Which province/territory do you think would provide the greatest opportunity for an immigrant? **v)** In what way do you think that immigration benefits the general Canadian population?

3. i) Answers may vary; for example:

a) *Identify problem/question* – does the addition of the HOV (Higher Occupancy Vehicle) lane really reduce congestion on the 401?

b) *Hypothesis* – the HOV lane does not reduce congestion on the 401 because people are still driving by themselves to work each day. (*added to help students focus on which methods would help them with their study)*

c) *Collect data* - interview, survey, experiment, case study or observation.

(i.e. interview – target regular commuters only, giving them a multiple-choice questionnaire that allows them to choose a pre-determined time spent commuting each day prior to and after the addition of the HOV lane, and limit the number of respondents to 100 sample with random sampling from the GTA (Greater Toronto Area))

d) *Analyze the data* – use bar graph to organize results and determine the degree of distribution

e) *Conclusion* – determine if data supports the hypothesis and answers the original question

ii) Answers may vary slightly; for an experiment or observation, peak times would be morning and evening rush hour, i.e. 6am-9am and 4pm-7pm. For an interview, contact commuters between 7pm and 8pm.

iii) Answers may vary; for example: interview - the interviewee would be restricted to select a pre-determined response (i.e. multiple-choice). Responses would be tabulated in a spreadsheet and then organized into a bar graph.

iv) Answers may vary slightly; for example: bar graph – compare height of bars 'prior to' and 'after' addition of HOV lane, using the properties of the distribution that may indicate if there are any relationships/correlations

(*emphasize that the sampling must ensure no bias*).

Unit 5 Day 5:Statistics – Collecting and Organizing One Variable Data			MBF 3C
	Description Collecting and Organizing One-Variable Data Collect, organize and store data from primary sources using appropriate sampling techniques.		Materials Internet, Excel/Fathom or Data Printouts BLM5.5.1,5.5.2 Assignment
		Assessment Opportunities	
Minds On…	**Whole Class → Discussion** Ask them how did the creation of their surveys go? 1. Take 5 minutes to proofread/ complete survey from the homework. 2. Pass out survey to a classmate. Record any possible sources of bias on the bottom of the survey.		This will likely need to be done in pairs or small groups to account for absences and incomplete work.
Action!	**Whole Class → Teacher Directed** **Part 1: Collect Data** Each student can access data on their chosen survey topic. Possible websites include: a) professional athlete earnings-- http://sportsillustrated.cnn.com/ b) top movie songs—www.afi.com c) immigration—www.statcan.ca d) cancer rates—www.cancer.ca OR provide handout of statistics. (BLM5.5.2) **Part 2: Assignment** 1. Import data into Excel/Fathom from the internet or create a graph by hand. 2. State each of the following: a) type of graph used (histogram, bar graph, circle graph, pictograph) and explain choice. b) distribution of your graph, if applicable (normal, bimodal etc.) c) whether the data came from a census or sample, justify your reasoning. d) 3 pieces of information about your data trends (i.e. conclusions that can be reached based on your graph)		
Consolidate Debrief	**Whole Class → Discussion** Review BLM **5.5.1 Notes** either as overhead or use examples to discuss proper survey design. Collect assignment		
Exploration	**Home Activity or Further Classroom Consolidation**		

MBF3C
BLM5.5.1

Notes

Example of survey on immigration:

1. From which continent did you originate?

□ South America □ Europe □ Asia □ Africa □ other

2. What do you believe is the main reason for which immigrants leave their place of origin?

□ war □ poverty □ lack of employment in specific field of training □ abuse □ other

3. What do you believe is the main reason for which immigrants come to Canada?

□ family □ climate □ employment opportunities □ democratic government □ other

4. Which province/territory do you think would provide the greatest opportunity for an immigrant?

□ British Columbia □ Alberta □ Ontario □ Quebec □ other

5. In what way do you think that immigration benefits the general Canadian population?

□ fills the absence of trained professionals □ able to learn new language □ able to try new foods □ increases cultural tolerance □ other

Please note:
Students will collect and analyze data for each question in survey. Emphasize to students that their survey must be able to produce some form of data that may be analyzed with a graph, etc. Selections within the survey may be perceived biased, but "other" is provided to allow for all viewpoints to be indirectly addressed

MBF3C

BLM5.5.2

Statistic Charts

a) Golf Player Earnings

Golf Player Earnings			
Rank	Player	Events	Money
1	Tiger Woods	10	$4,263,563.00
2	Jim Furyk	17	$4,174,516.00
3	Phil Mickelson	16	$4,123,005.00
4	Geoff Ogilvy	16	$4,003,049.00
5	Vijay Singh	17	$3,328,970.00
6	Trevor Immelman	17	$3,030,746.00
7	Stuart Appleby	16	$2,903,211.00
8	Adam Scott	13	$2,712,183.00
9	Chad Campbell	18	$2,424,507.00
10	Rory Sabbatini	17	$2,411,584.00
11	David Toms	15	$2,400,544.00
12	Carl Pettersson	20	$2,372,482.00
13	Stephen Ames	16	$2,227,035.00
14	Luke Donald	14	$2,188,642.00
15	Retief Goosen	12	$2,117,378.00
16	Brett Wetterich	16	$2,117,006.00
17	Rod Pampling	17	$2,092,767.00
18	Zach Johnson	19	$2,064,268.00
19	Jose Maria Olazabal	14	$1,953,102.00

Source: http://sportsillustrated.cnn.com/golf/pga/2006/stats/moneyleaders/

MBF3C

BLM5.5.2

Statistic Charts (continued)

Nascar Nextel Cup Driver Earnings

Rank	Player	Money Won
1	Jimmie Johnson	$5,959,217
2	Matt Kenseth	3,922,738
3	Jeff Burton	2,859,067
4	Kyle Busch	2,873,403
5	Kevin Harvick	3,396,052
6	Mark Martin	2,500,953
7	Kasey Kahne	3,717,698
8	Denny Hamlin	2,654,297
9	Jeff Gordon	3,599,247
10	Tony Stewart	4,047,555
11	Dale Earnhardt Jr.	3,124,598
12	Greg Biffle	2,707,056
13	Kurt Busch	2,942,046
14	Carl Edwards	2,673,001
15	Casey Mears	3,392,785

b) AFI top movie songs

#	SONG	MOVIE	YEAR
1	**Over the Rainbow** PERFORMER Judy Garland	WIZARD OF OZ, THE	1939
2	**As Time Goes By** PERFORMER Dooley Wilson	CASABLANCA	1942
3	**Singin' in the Rain** PERFORMER Gene Kelly	SINGIN' IN THE RAIN	1952
4	**Moon River** PERFORMER Audrey Hepburn	BREAKFAST AT TIFFANY'S	1961
5	**White Christmas** PERFORMER Bing Crosby	HOLIDAY INN	1942
6	**Mrs. Robinson** PERFORMERS Paul Simon, Art Garfunkel	GRADUATE, THE	1967
7	**When You Wish Upon A Star** PERFORMER Cliff Edwards	PINOCCHIO	1940
8	**Way We Were, The** PERFORMER Barbra Streisand	THE WAY WE WERE	1973
9	**Stayin' Alive** PERFORMER The Bee Gees	SATURDAY NIGHT FEVER	1977
10	**Sound of Music, The** PERFORMER Julie Andrews	SOUND OF MUSIC, THE	1965
11	**Man That Got Away, The** PERFORMER Judy Garland MUSIC/LYRICS	STAR IS BORN, A	1954
12	**Diamonds Are a Girl's Best Friend** PERFORMER Marilyn Monroe	GENTLEMEN PREFER BLONDES	1953
13	**People** PERFORMER Barbra Streisand	FUNNY GIRL	1968
14	**My Heart Will Go On** PERFORMER Céline Dion	TITANIC	1997
15	**Cheek to Cheek** PERFORMERS Fred Astaire, Ginger Rogers	TOP HAT	1935
16	**Evergreen (Love Theme from A Star is Born)** PERFORMER Barbra Streisand	STAR IS BORN, A	1976

Statistic Charts (continued)

AFI top movie songs (continued)

#	SONG	MOVIE	YEAR
17	**I Could Have Danced All Night** PERFORMER Audrey Hepburn (voiced by Marni Nixon)	MY FAIR LADY	1964
18	**Cabaret** PERFORMER Liza Minnelli	CABARET	1972
19	**Some Day My Prince Will Come** PERFORMER Adriana Caselotti	SNOW WHITE AND THE SEVEN DWARFS	1937
20	**Somewhere** PERFORMERS Natalie Wood (voiced by Marni Nixon), Richard Beymer (voiced by Jimmy Bryant)	WEST SIDE STORY	1961
21	**Jailhouse Rock** PERFORMER Elvis Presley	JAILHOUSE ROCK	1957
22	**Everybody's Talkin'** PERFORMER Harry Nilsson	MIDNIGHT COWBOY	1969
23	**Raindrops Keep Fallin' on My Head** PERFORMER B. J. Thomas	BUTCH CASSIDY AND THE SUNDANCE KID	1969
24	**Ol' Man River** PERFORMER Paul Robeson	SHOW BOAT	1936
25	**High Noon (Do Not Forsake Me, Oh My Darlin)** PERFORMER Tex Ritter	HIGH NOON	1952
26	**Trolley Song, The** PERFORMER Judy Garland	MEET ME IN ST. LOUIS	1944
27	**Unchained Melody** PERFORMER The Righteous Brothers	GHOST	1990
28	**Some Enchanted Evening** PERFORMER Rossano Brazzi (voiced by Giorgio Tozzi)	SOUTH PACIFIC	1958
29	**Born To Be Wild** PERFORMER Steppenwolf	EASY RIDER	1969
30	**Stormy Weather** PERFORMER Lena Horne	STORMY WEATHER	1943
31	**Theme from New York, New York** PERFORMER Liza Minnelli	NEW YORK, NEW YORK	1977
32	**I Got Rhythm** PERFORMER Gene Kelly	AMERICAN IN PARIS, AN	1951
33	**Aquarius** PERFORMERS Ren Woods, Ensemble	HAIR	1979
34	**Let's Call the Whole Thing Off** PERFORMERS Fred Astaire, Ginger Rogers	SHALL WE DANCE	1937
35	**America** PERFORMERS Rita Moreno, George Chakiris, Ensemble	WEST SIDE STORY	1961
36	**Supercalifragilisticexpialidocious** PERFORMERS Julie Andrews, Dick Van Dyke, Ensemble	MARY POPPINS	1964
37	**Swinging on a Star** PERFORMER Bing Crosby	GOING MY WAY	1944
38	**Theme from Shaft** PERFORMERS Isaac Hayes, Chorus	SHAFT	1971
39	**Days of Wine and Roses** PERFORMER Chorus	DAYS OF WINE AND ROSES	1963
40	**Fight the Power** PERFORMER Public Enemy	DO THE RIGHT THING	1989

BLM5.5.2

Statistic Charts (continued)

AFI top movie songs (continued)

#	SONG	MOVIE	YEAR
41	New York, New York PERFORMERS Gene Kelly, Frank Sinatra, Jules Munshin	ON THE TOWN	1949
42	Luck Be A Lady PERFORMERS Marlon Brando, Ensemble	GUYS AND DOLLS	1955
43	Way You Look Tonight, The PERFORMER Fred Astaire	SWING TIME	1936
44	Wind Beneath My Wings PERFORMER Bette Midler	BEACHES	1988
45	That's Entertainment PERFORMERS Fred Astaire, Nanette Fabray, Jack Buchanan, Oscar Levant	BAND WAGON, THE	1953
46	Don't Rain On My Parade PERFORMER Barbra Streisand	FUNNY GIRL	1968
47	Zip-a-Dee-Doo-Dah PERFORMER James Baskett	SONG OF THE SOUTH	1947
48	Whatever Will Be, Will Be (Que Sera, Sera) PERFORMER Doris Day	MAN WHO KNEW TOO MUCH, THE	1956
49	Make 'Em Laugh PERFORMER Donald O'Connor	SINGIN' IN THE RAIN	1952
50	Rock Around the Clock PERFORMERS Bill Haley and the Comets	BLACKBOARD JUNGLE	1955

Unit 5 Day 6: Statistics – Measures of Central Tendency

MBF 3C

Description

Measures of Central Tendency
Calculating mean, median and mode and identifying when each is a good choice.

Materials

Graphing Calculators

Assessment Opportunities

Minds On…

Whole Class → Two Corners

Read the following scenario and have the students stand on either Rahim's side of the room or Johann's side of the room depending on who they feel is a better candidate. Ask them to reflect upon why their answer seems reasonable.

Two car salesmen are competing for a mid-year bonus. The owner of the dealership wants to assess the better competitor. Who is the better candidate?

Monthly Sales

Rahim	16	28	32	28	26	31
Johann	34	30	24	26	29	26

This depends on how the owner judges the centre of their data.

Things for the teacher to bring up once their arguments have subsided.

If she wants to look at the traditional average, then Rahim's is $\frac{161}{6} = 26.8$. This means that he sells an average of 26.8 cars a month. Similarly, Johann's is $\frac{169}{6} = 28.2$. By this calculation, Johann sells more cars a month.

However, if we look more closely at their data, Rahim is more likely to sell 28 cars in a month and Johann is more likely to sell only 26 cars in a month, because these are their middle number of sales.

This set of data explains why it is important to do as many calculations as possible before summarize a set of data.

Possible discussion on meaning of average: lead into mean, vs. median, vs. mode.

Action!

Whole Class → Direct Instruction

Measures of Central Tendency Lesson:

*There are 3 ways to find the common trend (or central tendency) for a set of data.

1) Mean (most commonly referred to as the average)
*To find the mean, add up all of the numbers in your list and divide by the number of numbers.

Ex 1: Jesara is buying a home that will require a mortgage. The bank wants to know her monthly salary. She works on commission, so she must calculate her average salary. Given her income for the first 6 months of the year, what is her average salary?

Jan--$3675, Feb--$4250, Mar--$3225, Apr--$2985, May--$3650, Jun--$4600.

Solution: Mean

$$= \frac{3675 + 4250 + 3225 + 2985 + 3650 + 4600}{6}$$

$$= \frac{22385}{6}$$

$$= \$3730.83$$

$\therefore$ She would tell the bank that she makes an average of \$3730.83/ month.

2) Median

*The median is the middle entry in an ordered list. There are as many data points above it as below it.

*To find the median,

a) If there is an odd number of data points, take the middle one (i.e. if there are 13 numbers, the median is the value of the 7^{th} number when they are listed in ascending order).

b) If there is an even number of data points, the median is the average of the middle two numbers.

Ex 2: Find the median mark for each list of student grades.

a) 62, 64, 76, 89, 72, 54, 93

First, list the numbers in ascending order.

54, 62, 64, 72, 76, 89, 93

^-- one middle number

med = 4^{th} entry

= 72

b) 56, 84, 63, 67, 62, 98

56, 62, 63, 67, 84, 98

^--^ two middle values

med= average of 3^{rd} and 4^{th}

=(63+67) ÷ 2

= 65

3) Mode

*The mode is the most frequent number in a data set.

*There can be no mode as well as more than one mode.

Ex 3: Find the mode(s) for each list of numbers.

a) 5, 7, 9, 8, 6, 5, 4, 10

mode= 5

b) 25, 30, 32, 30, 25, 29

modes= 25 and 30

c) 63, 57, 66, 83, 79, 72, 79, 69, 60, 63, 79, 85, 80

mode= 79

Ex 4: The modes of the following set of data are 7 and 9. What must be the value of *y*?

6, 9, 3, 4, 8, 0, 7, 2, 9, *y*

Sol'n: Since both 7 and 9 must be there the same number of times, *y* must be 7.

Note: Both of the **mean** and **median** calculations can be done on the graphing calculator. The advantage is that you don't need to enter the data in ascending order.

		Steps for the graphing calculator: 1. Enter the data into L_1 by pressing **STAT 1:EDIT** 2. Press **STAT** and scroll over to **CALC** 3. Press **1** for 1-Var Stats 4. Type $\mathbf{L_1}$ by pressing **2^{nd} 1 ENTER** 5. The mean is given by $\bar{x}$. The median is found by scrolling down the list past the original screen to the word *med*. TI-89 Titanium Instructions 1. Go to Stats/List Editor Program from the APPS list 2. Enter the data into List 1 3. Press F4 button → Calc 4. Select 1 → 1-Var Stats 5. For the List data box you need to move the cursor into that box and press 2^{nd} button and then the minus button (VAR-LINK) then scroll down until you find "list 1" which you then select and press enter (If you were using two lists where list 1 contained the data values and list 2 contained the frequencies – you would in the "Freq" box follow the same instructions as above to place "list 2" there) 6. Press enter to accept the settings 7. Wait for a few moments for the calculations to occur and then you can scroll up and down through the values. $\bar{x}$ is the mean and MedX is the median. **Ex 5:** Using the graphing calculator, find the mean, median and mode for the heights of 15 rugby players. 182, 178, 181, 182, 172, 176, 183, 177, 173, 176, 185, 181, 177, 182, 175 mean= 178.7 med= 178 mode= 182 (done by hand)		
	Consolidate Debrief	**Small Groups → Discussion** Ask the students if they can come up with "tips" of when to use mean, median and mode. *All 3 measures of central tendency are good indicators of the trend in data, but at times some are better choices than others. ***Tips:*** Mean—Really good when the data is fairly close together. Most commonly used. Median—Good when there is an outlier (i.e. a number that is far away from the others which would skew the mean). Mode—Good when the value of the number is the most important information (e.g. shoe size). --Only choice with categorical data.		
Concept Practice		**Home Activity or Further Classroom Consolidation** Students complete BLM 5.6.1		The graphing calculator can be used for any of the homework questions.

MBF3C Name:

BLM 5.6.1 **Measures of Central Tendency** Date:

1. Find the mean, median and mode for each set of data.
   a) 64, 69, 72, 54, 89, 92, 54, 32 b) 12, 0, 8, 4, 6, 3, 7, 3, 2, 9, 5, 6, 7, 7, 8
   c) 4.2, 11.5, 6.8, 5.2, 5.4, 6.3, 12.1, 11.5, 11.9, 7.8, 13.1, 5.8, 6.2
   d) 0, 2.1, 5.7, 8.3, 2.6, 7.3, 8.4, 0.5, 0.4, 2.1, 2.2, 4.3, 5.7
   e) $\frac{3}{4}, \frac{5}{6}, \frac{8}{3}, \frac{7}{2}, \frac{1}{6}$ (answer in fraction form)

2. Gabriel buys 8 DVDs at Discount Dan's DVD shop. Three cost $10.50, 2 cost $7.75, 1 cost $5.25 and 2 cost $3.50. Find the mean, median and mode of the costs of his DVDs.

3. The prizes in the local lottery were worth the following:
   2 prizes of $1 000 000
   7 prizes of $350 000
   10 prizes of $250
   Find the mean, median and mode.

4. The masses, in kilograms, of group of Jessy Bragg's weight loss group are shown.
   81, 79, 83, 76, 89, 75, 67, 83, 65, 74, 78

   a) Find the mean, median and mode.
   b) Is the median greater than or less than the mean?
   c) Is the mode greater than or less than the mean?

5. The hourly rates of employees of a supermarket are given.
   $9.25, $8.50, $22.50, $7.85, $8.85, $12.65, $10.85, $11.50

   a) Find the mean, median and mode.
   b) Which of your answers best represents the data? Why?
   c) Which of your answers would most misrepresent the data? Why?

6. a) Find the mean and median for each set of marks.
   Suzy: 25, 36, 39, 87, 89, 94
   Ruiz: 45, 56, 88, 89, 92, 98

   b) What is the best measure of central tendency for Suzy, the mean or the median?
   c) What is the best measure of central tendency for Ruiz, the mean or the median?

7. State and explain whether each statement is based on the mean, median or mode.
   a) 0.2% of light bulbs are defective.
   b) The most popular search engine is Google.
   c) The average university grad earns $35 000 annually upon graduation.
   d) Most drinking and driving accidents occur on long weekends.

8. You earned the following marks (each out of 50) on your first five test: 28, 36, 38, 41, 44. What mark would you have to get on the next test in order to bring your test average up to 80%?

MBF3C
BLM 5.6.1

Measures of Central Tendency Solutions

1. a) mean = 65.8; median = 66.5; mode = 54 **b)** mean = 5.8; median = 6; mode = 7 **c)** mean = 8.3; median = 6.8; mode = 11.5 **d)** mean = 3.8; median = 2.6; modes = 2.1 and 5.7 **e)** mean = 95/12; median = 5/6; mode = n/a
2. mean = $6.58; median = $7.75; mode = $10.50 **3.** mean = $234 342.10; median = $250; mode = $250 **4. a)** mean = 77.3; median = 78; mode = 83 **b)** greater than **c)** greater than **5. a)** mean = $11.49; median = $10.05; mode =n/a **b)** median; wide range of wages **c)** mean; data is not close together
6. a) Suzy: mean = 61.7; median = 63; Ruiz: mean = 78; median = 88.5 **b)** median **c)** median **7. a)** mean **b)** mode **c)** median (grads make a wide range of salaries depending on their degree/subject area) **d)** mode
8. a) 106% or 53 out of 50.

Unit 5 Day 7: Statistics - Standard Deviation		MBF 3C
	Description Measures of Spread Calculate and interpret range and standard deviation by hand and with technology.	Materials Graphing Calculators BLM 5.7.1

Assessment Opportunities

Minds On…

Whole Class→ Discussion

What can you infer, justify and conclude about the Joaquin's and Taran's tests scores? (Hint: Calculate the mean, median and mode for each. What do they tell you?)

Joaquin's Tests: 76, 45, 83, 68, 64
Taran's Tests: 67, 70, 70, 62, 62

J.'s mean= 67.2
med= 68
mode= none

T.'s mean= 66.2
med= 67
mode= 62, 70

This tells us that Joaquin has a higher average. Still, by looking at the data, we can see that Taran is more consistent.

Action!

Whole Class → Teacher Directed

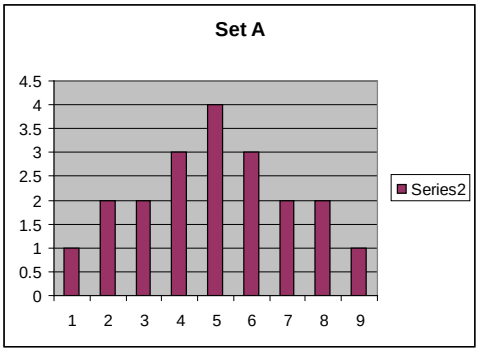
Measures of Spread Lesson:

*Mean, median and mode are all good ways to find the centre of your data.
*This information is most useful when the sets of data being compared are similar.
*It is also important to find out how much your data is spread out. This gives a lot more insight to data sets that vary from each other.

Ex 1: Consider the following two data sets with identical mean and median values. Why is this information misleading?

Set A) 0, 2, 2, 4, 4, 6, 6, 6, 8, 8, 8, 8, 10, 10, 10, 12, 12, 14, 14, 16

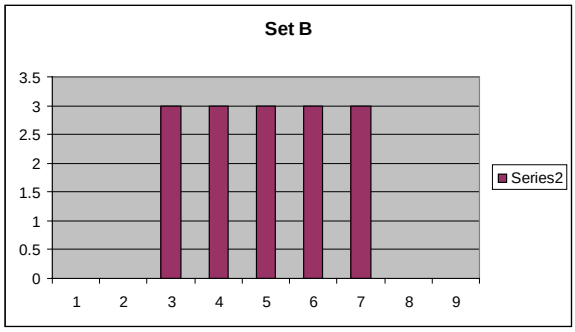
mean= 8
med= 8



Set B) 4, 4, 4, 6, 6, 6, 8, 8, 8, 10, 10, 10, 12, 12, 12,

mean= 8
med= 8

Opportunity to talk about the real-life implications of consistency in data. i.e. the more consistent the data, the more significant it becomes, and the more it is cited in the media etc.



Sol'n: This information is misleading because one graph is bell-shaped and the other is uniform, but the calculations make them appear to be similar when really A and B are spread out quite differently.

What is something that can be done to further compare these graphs?

Look at the ***range*** in the data sets.

Range—the difference between the highest and lowest numbers.

A) Range = 16-0
= 16

B) Range= 12-4
=8

$\therefore$ Set B is more consistent since it has a smaller range.

Ex 2: Twins, Toby and Moby, both work at a local pizza shop. Their manager has decided to give a raise to her best employee. She looks at their data.

Number of Pizzas Made per Shift

Toby	54	152	180	12	72	126	104	132
Moby	132	104	102	120	86	12	180	96

Who is more deserving?

Sol'n: She starts by finding the mean number of pizzas made by each and their range.

Toby: $mean = \frac{832}{8}$
$= 140$
range = 180-12
=168

Moby: $mean = \frac{832}{8}$
$= 140$
range = 180-12
=168

These statistics leave both employees equal.
The manager notices that Moby's data looks more consistent, but she needs proof to support her claim.
She decides to calculate the ***standard deviation*** for each.

Standard Deviation (σ)—best choice for measuring the spread of data

Steps for calculating σ:

1. Find the difference between each value and the mean.
2. Square each difference.
3. Add up all of your answers from Step 2.
4. Divide this sum by the number of numbers (i.e. find the average of the differences squared).
5. Find the square root your answer.

Mathematically: $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}}$

where σ = standard deviation
$\bar{x}$ = mean
n = number of entries

Standard deviation for Toby (by hand):

Number of Pizzas x	$x - \bar{x}$	$\left(x - \bar{x}\right)^2$
54	54-140=-86	7396
152	12	144
180	40	1600
12	-128	16384
72	-68	4624
126	-14	196
104	-36	1296
132	-8	64
	Total=	31704

$$\sigma = \sqrt{\frac{31704}{8}}$$
$$\sigma = \sqrt{3963}$$
$$\sigma = 62.95$$

In order for this standard deviation to be significant, you must compare it to another data set.

Standard deviation for Moby (with the graphing calculator):

Steps:
1. Enter the data into L_1 by pressing **STAT 1:EDIT**
2. Press **STAT** and scroll over to **CALC**
3. Press **1** for 1-Var Stats
4. Type $\mathbf{L_1}$ by pressing **2^{nd} 1 ENTER**

Note: for the TI-89 Titanium calculator use the same steps for the mean/median calculations from last day and as stated below look for the value given by σx**.**

The standard deviation is given by σx**.**
$\therefore$ the standard deviation for Moby is 64.54

$\therefore$ Toby's σ is smaller.
$\therefore$ Toby's data is closer to the mean than Moby's.
$\therefore$ Toby is more consistent and deserves the raise.

		Ex 3: Find the range and standard deviation of the following set of numbers: 3, 10, 8, 20, 4, 4, 3, 8, 8, 8, 12 **Sol'n:** Range= Highest Value- Lowest Value = 20-3 =17 Standard Deviation (on graphing calculator) = 4.73		
	Consolidate Debrief	**Whole Class→ Discussion** Revisit the problem from Minds On and now take a look at the range and standard deviation. What can you infer, justify and conclude about the Joaquin's and Taran's tests scores? Joaquin's Tests: 76, 45, 83, 68, 64 Taran's Tests: 67, 70, 70, 62, 62		
Application *Concept Practice*		**Home Activity or Further Classroom Consolidation** Students complete BLM 5.7.1		

MBF3C Name:

BLM 5.7.1 **Standard Deviation** Date:

1. True or False? The standard deviation cannot be negative.

2. Calculate the range and standard deviation of each.
   a) 4, 8, 6, 3, 12, 9, 7, 6 b) 35, 38, 40, 43, 46, 23, 38
   c) 2.4, 4.3, 6.5, 1.1, 8.9, 3.6, 7.2, 9.6 d) 4.55, 3.23, 6.78, 3.54, 5.54, 6.78

3. The machine packaging cookies has been considered defective. The packages are labelled as containing 150g. A sample of 15 packages was selected and the masses are given.

   145, 151, 152, 150, 147, 152, 149, 148, 153, 150, 146, 152, 148, 149, 151

   a) Calculate the mean.
   b) If any packages are more than 2.2g from the mean, the package is not sold. How many are defective?
   c) Should the machine be fixed?

4. A group of student landscapers are to keep track of their own weekly hours. They are listed below.
   44, 52, 43, 39, 42, 41, 38, 43, 46, 45, 44, 39, 40, 42, 45

   a) Find the range. Is this a useful tool for representing this data?
   b) Find the mean.
   c) Find the standard deviation.
   d) What can be said about the entry of 52 hours/week?
   e) Calculate the standard deviation again without the 52 hours/week entry.

5. The sale prices of the last 10 homes sold in 1985 were: $198 000, $185 000, $205 200, $225 300, $206 700, $201 850, $200 000, $189 000, $192 100, $200 400.
   a) What is the average sale price?
   b) What is the standard deviation?
   c) Do you think that a price of $240 000 would be considered unusual? Why or why not?

6. The sales price of the last 10 homes sold in 2005 were: $345 500, $467 800, $289 000, $675 000, $398 500, $243 000, $899 950, $453 000, $239 000, $256 000.
   a) What is the average sales price?
   b) What is the standard deviation?
   c) Which year was more consistent? How do you know?

Solutions

1. true **2. a)** range = 9; s.d. = 2.85 **b)** range = 23; s.d. = 7.37 **c)** range = 8.5; s.d. = 3.08 **d)** range = 3.55; s.d. = 1.55 **3. a)** 149.5 **b)** 7 **4. a)** 14; yes, to ensure that students are being honest when recording the number of hours worked **b)** 42.9 **c)** 3.50 **d)** answers may vary; for example: over-inflated **e)** 2.52 **5. a)** $200 355 **b)** 11 189.04 **c)** yes, based on the standard deviation, it would be an extremely high value **6. a)** $426 675
b) 214 078.1 **c)** 1985; smaller range of values

Unit 5 Day 8: Statistics - Analyzing One Variable Data

MBF 3C

Description

Analyzing One-Variable Data

Compare 2 sets of data using central tendency and measures of spread. Solve problems by interpreting and analyzing one-variable data from secondary sources.

Materials

Internet/ copied data sets, graphing calculators/ Excel, Investigation BLM5.8.1

Assessment Opportunities

Minds On…

Pairs → Review

Find the range and the standard deviation of : 2, 5, 10, 6, 4, 3, 4, 6.

Sol'n:

Range= 10-2

= 8

Standard Deviation:

x	$x-\bar{x}$	$(x-\bar{x})^2$
2	2-5= -3	9
3	-2	4
4	-1	1
4	-1	1
5	0	0
6	1	1
6	1	1
10	5	25

$$\sigma = \sqrt{\frac{42}{8}}$$

$$\sigma = 2.29$$

Action!

Whole Class → Teacher Directed

Analyzing One-Variable Data Lesson:

Consider the following list of data collected from MLS:

Average Home Prices by Province (in $)

Province	June 2006	June 2005
Yukon	177191	159668
Northwest Territories	243745	250222
British Columbia	399829	330333
Alberta	294282	215964
Saskatchewan	134161	121984
Manitoba	155531	139195
Ontario	280263	268074
Quebec	284747	252745
New Brunswick	127406	123732
Nova Scotia	170547	157524
Prince Edward Island	134115	114223
Newfoundland/ Labrador	132571	140958

The first part of lesson could be done together as an example for how the investigation should go.

		Compare the two years by finding all measures of central tendency and measures of spread. ***Mean:*** **2006** = $211 199 **2005** =$189 551.80 ***Median:*** **2006** =$173 869 **2005** =$158 596 ***Mode:*** **2006** None **2005** None ***Conclusions:*** On average, housing prices rose $21 647.20 during this year. However, the median only changed by $15 273. This indicates that there was at least one outlier which had a greater change than the other provinces. Looking at the chart, BC and Alberta are both outliers. Now, it is important to examine the data's accuracy. ***Range:*** **2006** =$399 829- 127 406 =$272 423 **2005** =$330 333- 114 223 =$216 110 ***Standard Deviation:*** **2006** =$87 755.31 **2005** =$71 187.98 ***Conclusions:*** The 2006 values are much less accurate since they have both a larger range and significantly larger standard deviation. This indicates that there is much less consistency in 2006. **Part Two: Investigation 5.8** Students can work to complete the same question for a different set of data. See handout entitled ***Investigation BLM5.8.1***.		All of these calculations were done with the use of Excel. The graphing calculator would be just as good. Possible discussion: Can the 2006 data be accurately used to forecast future trends? Any other 2 sets of data would work well here.
	Consolidate Debrief	**Whole Class → Discussion** As a classroom group, discuss the amount of the investigation that is still incomplete. Note: The graph can be done by hand or on Excel. Have students focus on the conclusions given by their calculations.		
Exploration Reflection		**Home Activity or Further Classroom Consolidation** 1. Complete the investigation. 2. Start on a unit summary sheet (mind map).		

MBF3C Name:

BLM 5.8.1

Investigation

Date:

All questions can be done by hand or with the use of technology.

Compare *any two* of the following sets of data. Be specific when giving conclusions.

Canadian Housing Prices by City ($)

Canadian City	June 2006	June 2005	2002
Vancouver	508 435	422 843	301 473
Victoria	538 913	469 588	242 503
Calgary	367 033	245 803	198 350
Edmonton	254 240	199 409	150 165
Regina	137 022	132 054	100 751
Saskatoon	160 548	139 728	118 999
Ottawa	260 458	254 725	200 711
Toronto	358 035	345 065	275 975
Montreal	222 879	210 740	143 589
Fredericton	136 371	134 334	114 185
Saint John	127 586	125 455	104 052
Halifax	201 316	184 853	148 737

Sources: MLS and Remax

1. State the two sets of data that you wish to compare here: _____

2. Find the measures of central tendency for each. State any conclusions found.

Mean: **1st Set** **2nd Set**

Median: **1st Set** **2nd Set**

Mode: **1st Set** **2nd Set**

Conclusions: _____

MBF3C Name:

BLM 5.8.1

Investigation (continued)

Date:

3. Find the measures of spread for each. State any conclusions found.

Range: **1st Set** **2nd Set**

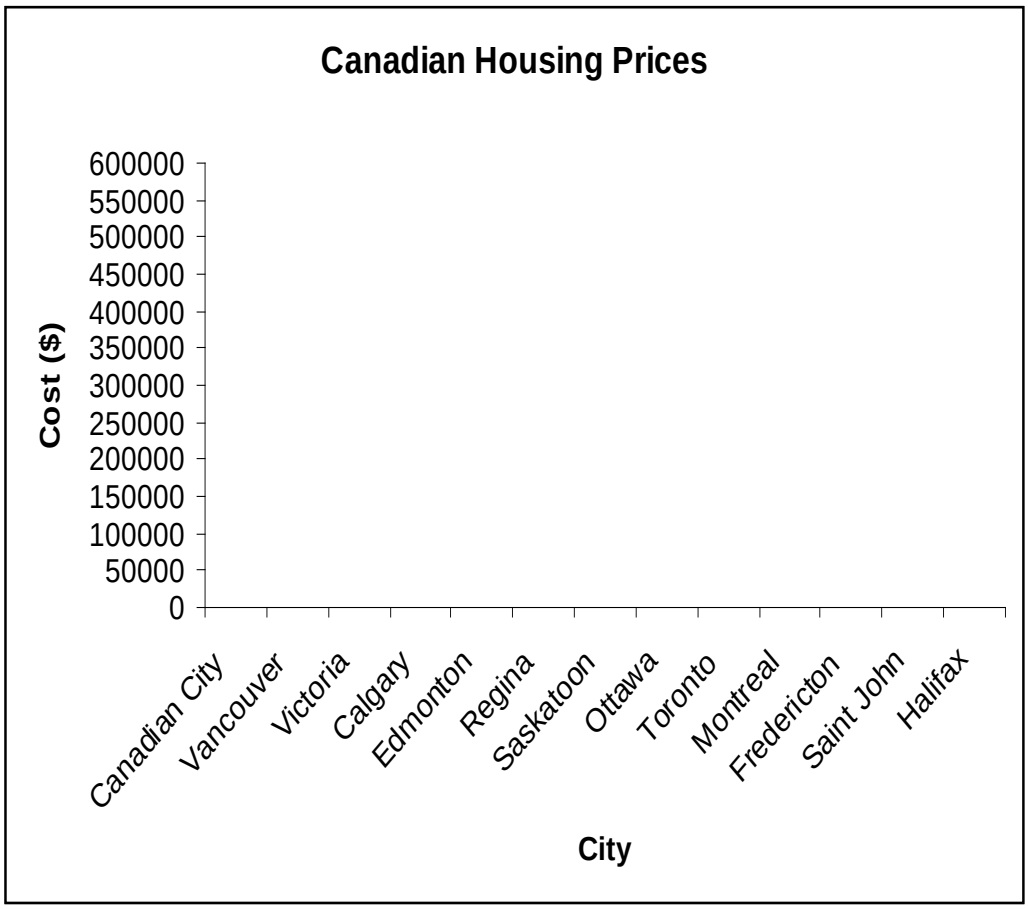
Standard Deviation: **1st Set** **2nd Set**

Conclusions: _____

_____

_____

4. Graph your data.



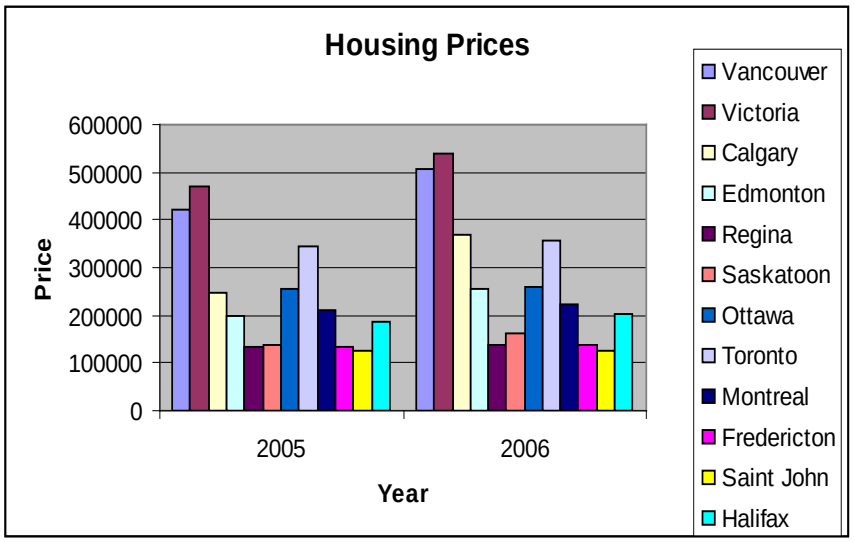
MBF3C

BLM 5.8.1

Investigation Solutions

2006 vs. 2005

Canadian City	Jun-06	Jun-05
Vancouver	508435	422843
Victoria	538913	469588
Calgary	367033	245803
Edmonton	254240	199409
Regina	137022	132054
Saskatoon	160548	139728
Ottawa	260458	254725
Toronto	358035	345065
Montreal	222879	210740
Fredericton	136371	134334
Saint John	127586	125455
Halifax	201316	184853
Mean	272736.3	238716.4
Median	238559.5	205074.5
Range	411327	344133
Standard Deviation	141844.9	116365.1



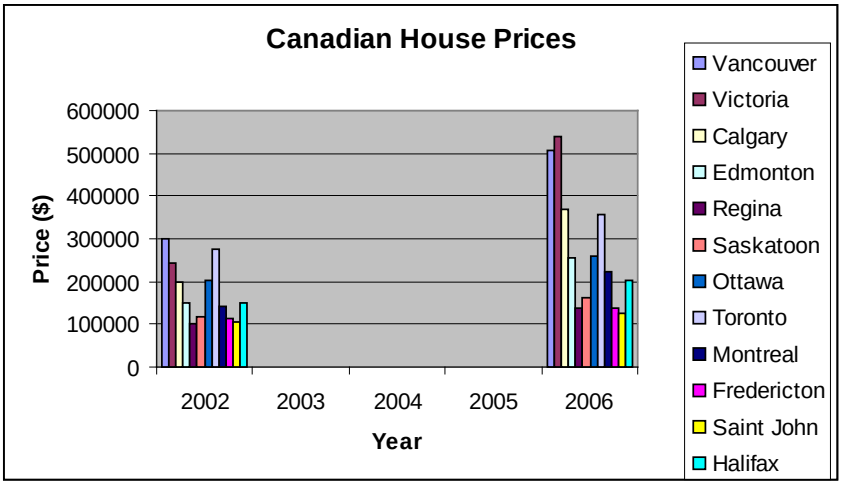
MBF3C

BLM 5.8.1

Investigation Solutions Continued

2006 vs. 2002

Canadian City	Jun-06	Jun-02
Vancouver	508435	301473
Victoria	538913	242503
Calgary	367033	198350
Edmonton	254240	150165
Regina	137022	100751
Saskatoon	160548	118999
Ottawa	260458	200711
Toronto	358035	275975
Montreal	222879	143589
Fredericton	136371	114185
Saint John	127586	104052
Halifax	201316	148737
Mean	272736.3	174957.5
Median	238559.5	149451
Range	411327	200722
Standard Deviation	141844.9	68509.17



MBF3C

BLM 5.8.1

Investigation Solutions continued

2005 vs. 2002

Canadian City	Jun-05	Jun-02
Vancouver	422843	301473
Victoria	469588	242503
Calgary	245803	198350
Edmonton	199409	150165
Regina	132054	100751
Saskatoon	139728	118999
Ottawa	254725	200711
Toronto	345065	275975
Montreal	210740	143589
Fredericton	134334	114185
Saint John	125455	104052
Halifax	184853	148737
Mean	238716.4	174957.5
Median	205074.5	149451
Range	344133	200722
Standard Deviation	116365.1	68509.17