Name: _____Ch 5 Organization of Data – 5.2/5.5A MDM 4U1

5.2 Principles of Data Collection

Suppose you want to a study or survey to gather information. There are a few ways you can do it.

1) You can ask every single person in the *population*. This would be considered a **census.** While the information it contains is reliable and valuable, it may difficult to perform, time-consuming and quite expensive.
   OR
2) You survey a small group selected from the overall population. This is called a **sample.**

A. Election time!

The 2018 Ontario Provincial Elections are approaching. Here are the candidates for the Etobicoke Centre riding, (in alphabetical order, by last name).

- Yvan Baker (Liberal)
- Erica Kelly (NDP)
- Shawn Rizvi (Green)
- Kinga Surma (PC)

Suppose you would like to know which candidate in the Etobicoke Centre riding is likely to receive the most votes. You decide to **sample** the BCI population.

With your class peers, try determining the kind of sample you might use, below.

a) **Where** and **to whom** would you give the survey?

b) **How many** surveys would you give out?

c) How similar would you expect the **results** from another MDM 4U1 group's survey to be the same or similar to yours? Explain.

Name: _____Ch 5 Organization of Data – 5.2/5.5A MDM 4U1

--stop here and check in with your teacher before continuing----

Each sample that is performed will be different from other samples. This is called

variability.

As the samples get similar to each other (and as the sample size increases), you expect the variability between samples *to decrease* and the sample to be more representative of the overall population.

If a sample is meant to be representative of the larger population, there are a few things to keep in mind when designing it:

- Selection must be *random*, so that each member has an equally likely chance of being selected. If it's not random, it is biased and unreliable.

There are many different kinds of samples that can be performed:

Type of Sample		**Example**
Simple Random – randomly choose a specific number of people. Includes systematic and stratified samples.		Put everyone's name into a hat and draw five names.
	Systematic – put the population into an ordered list, and choose people at regular intervals	Order everyone in some way (i.e. alphabetically) and select every 10^{th} name
	Stratified – divide the sample into groups with the same proportion as the groups have in the general population. (Time and cost efficient)	Out of 633 female and 367 male employees, 63 women and 37 men are randomly chosen
Cluster – divide the population into groups, randomly choose a number of the groups, sample each member in those groups		Of all the schools in TDSB, select 5 schools and ask all the staff at those schools.
Multistage – divide the population into a hierarchy, and choose a random sample at each level		Survey of a chain of department stores: randomly select 10 stores, then randomly select 3 departments, then randomly select 10 employees
Convenience – choose individuals from the population who are easy to access		A person goes to a local park and asks the nearby people their thoughts on parks and recreation
Voluntary – allow participants to choose whether or not to participate in the survey		Online poll asking if schools should ban junk food

Name: _____Ch 5 Organization of Data – 5.2/5.5A MDM 4U1

B. <u>Practice.</u> Classify each of the following types of samples. Explain your reasoning.

1. A reporter stops people on the street to ask for opinions on a proposed legal change.

2. A teacher wishes to get feedback from a class about a recent presentation. She puts the names of all 30 students into a hat and draws 5 names.

3. A company is interviewing its employees about whether they would use an in-house fitness centre. They plan to interview every 15^{th} person on the payroll list.

4. A market research company mails surveys to all households, and asks residents to mail back their response in prepaid envelopes.

5. The province wants to interview 250 randomly chosen students. It randomly selects 5 school boards, then randomly selects 5 schools, then randomly selects 10 students from each of those schools.

6. A grocery store wants to know if it should open a branch in Huntsville, Ontario. They decide to poll individuals from Huntsville and three neighbouring, smaller towns, by surveying 25% of the population in each town.

Check with teacher, then 5.2 HW pg 210 # 1, 2 4, 7, 8

Name: _____Ch 5 Organization of Data – 5.2/5.5A MDM 4U1

5.5A – Bias in Sampling

A. Types of Samples: Let's revisit the types of samples in 5.2.

1. What are some problems with **convenience** samples? Some benefits?

2. What are some problems with **voluntary** samples? Some benefits?

B. Explain the problem(s) with each of the following:

1. A teacher asks students to raise their hands if they cheated on last week's test.

2. A politician goes back to the farming community where they grew up to see if they support the latest initiatives in the agricultural industry.

3. A survey question asks: "A lot of people do not like math. How do you feel being referred to as a math geek?"

4. A mailed survey only has a 3% rate of mail-in returns for the survey.

--stop here and check in with your teacher before continuing----

All the examples from Part B, above, show a type of **bias,** or a prejudice either in favour or against.

Example	Type of Bias	Definition
A teacher asks students to raise their hands if they cheated on last week's test.	**Response Bias**	Respondents **change their answers** to influence results, to avoid embarrassment, or to give the answer they think the examiner wants Ex: Students who cheated choose to leave their hand down so they aren't embarrassed.
A politician goes back to the farming community where she grew up to see if they support the latest initiatives in the agricultural industry.	**Sampling Bias**	There is a problem with **who is sampled**. The sample does not closely represent the larger population. Ex: A higher-than-average proportion of the community consists of farmers, who are likely to support the initiative, since it benefits them and since the politician grew up there.
A survey question asks: "A lot of people do not like math. How do you feel being referred to as a math geek?"	**Measurement Bias**	There is a problem with the **wording of the sample itself**. Some characteristics are consistently over or under represented. Ex: The math question is a **leading question** that can influence the responder. Other measurement bias includes **limited/insufficient options** for M/C responses, etc.
A mailed survey only has a 3% rate of mail-in returns for the survey.	**Non-response Bias**	Occurs when many people don't respond, and the opinion of those who do respond is significantly different from the population. Typically, **those who respond have very strong opinions** (either

		positive or negative) that are not representative of the population.

C. Practice. Identify the type of bias in each sample, and explain your reasoning.

1. You ask only students on sports teams how to spend the school fundraising money.

2. A survey question asks, “Who is the best basketball player of all time, Michael Jordan or Lebron James?”

3. A radio call-in show asks callers to ask the question, “Are you in favour of the law that would require pet owners to register their pets?”

4. A survey question asks, “How many words per minute can you read?”

5. A phone company sends texts messages to customers and asks which services they like the best.

6. A survey asks, “How do you feel about the development of Light Rail Transit across Eglinton Ave.?” and has these response answers: “Strongly support; Mostly support; Somewhat support”

Name: _____Ch 5 Organization of Data – 5.2/5.5A MDM 4U1

Check with teacher, then 5.5A Bias HW pg 240 # 1 - 3