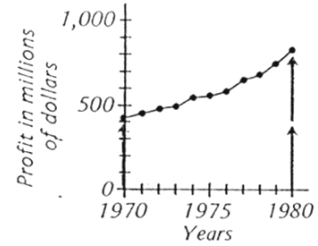
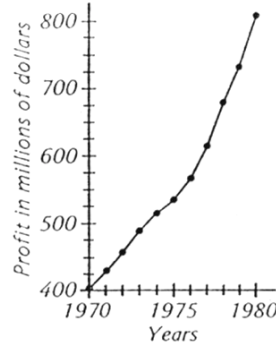
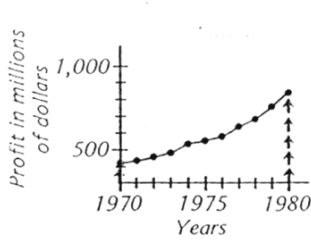


The Power of Data - The Media

Statistical information is often presented in graphical form. Unfortunately, graphs can also be used to deceive. There are ways in which a graph may be used to create a false impression of the data that it represents.

Ex. 1: Each of the graphs below represents the data found on the following table:

Year	Profits in millions
1970	403
1971	429
1972	458
1973	491
1974	515
1975	536
1976	574
1977	620
1978	677
1979	734
1980	811

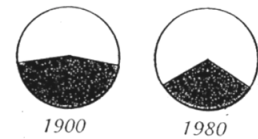


What are some observations about the three graphs as compared to the table?

Although graphs are used to present statistical information because of the simple picture that they can give of the patterns it contains, they can also be deliberately drawn to give a false impression. And even a graph that is meant to give an honest representation can be misinterpreted.

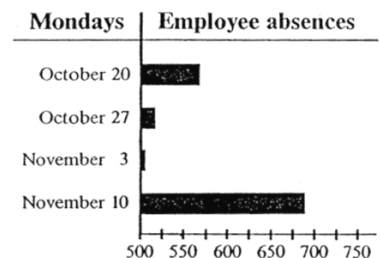
Ex. 2: The shaded slices of these circle (or “pie”) graphs show the parts of the population of the United States in 1900 and 1980 that were less than nineteen years old.

Part of the U.S. population less than nineteen years old



- a) In which year was a greater percentage of the population less than nineteen years old?
- b) Does it follow that there were more people in the United States less than nineteen years old in that year than the other? Explain.

Ex. 3: This graph appeared several years ago in a bulletin for employees of a large company. It compared the number of absences on four consecutive Mondays, the last preceding a legal holiday.

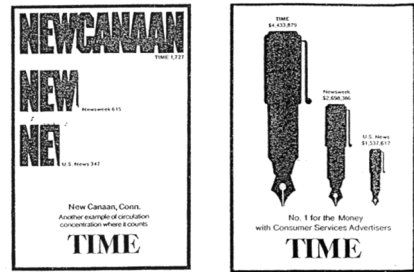


- a) What impression does the graph seem to intend to convey?

- b) The actual numbers of absences on these four days were 566, 517, 501, and 689. Do the lengths of the four bars make this obvious?

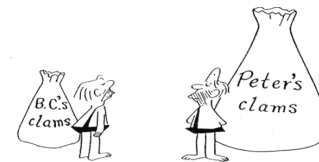
- c) Why does the graph give the impression that it does?

Ex. 4: Which pictograph is misleading? Why?



Picture graphs, frequently used to make comparisons, can be easily misinterpreted.

Ex. 5: Suppose that Peter has twice as many clams as B. C. and that we decide to show this with two bags of clams, one drawn twice as tall as the other.

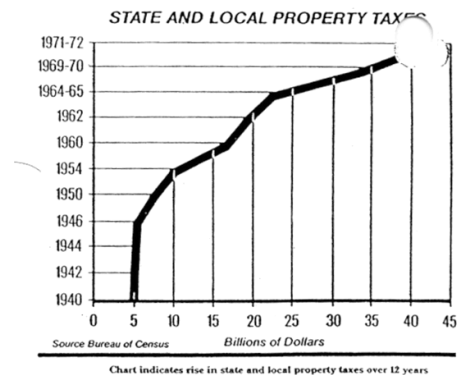


- a) Does the larger bag look as though it contains exactly twice as many clams as the smaller one?

- b) Rather than comparing the heights of the two bags, someone might compare the areas covered by them. How do you think their areas compare? *Hint:*

- c) If the number of clams that each bag contains is determined by its volume, how do you think their volumes compare? *Hint:*

Ex. 6: What's wrong with this graph?



Data Analysis with Graphs

A. Definitions

The study of statistics can be very rewarding given the patience and precision it deserves. To begin this unit, we'll first look at some statistical language that we need to understand before we continue.

Statistics –

Raw Data –

Datum –

Variable –

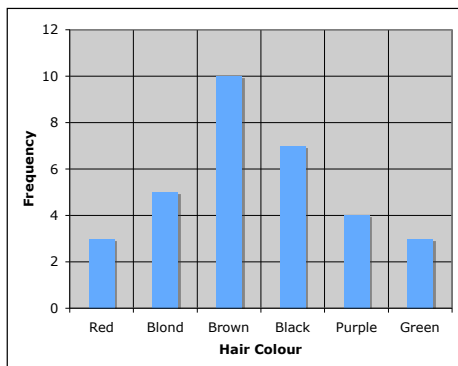
Continuous Variable –

Discrete Variable –

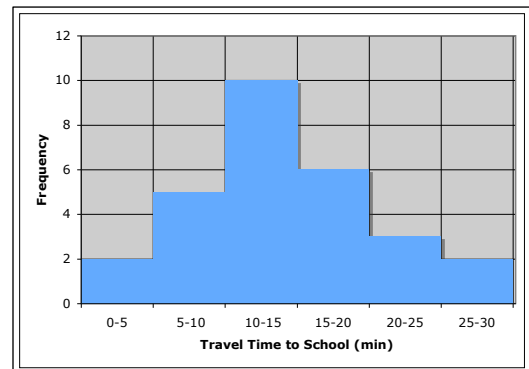
Interval (or Group) –

Range –

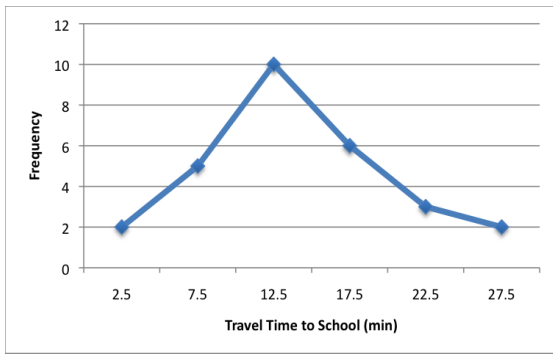
Bar Graph –



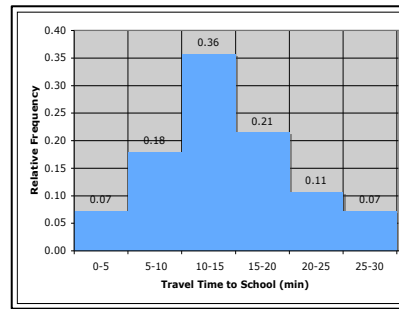
Histogram –



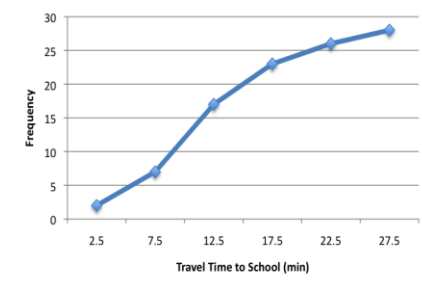
Polygon –



Relative Frequency Histogram –



Cumulative Frequency Polygon –



B. Examining Data and Constructing Graphs

When presented with a set of data, there are a number of questions we must pose before analysis can continue. One needs to consider the nature of the data, is it continuous or discrete? This will determine the type of graph we create. We should also ask if we should use a grouped or ungrouped format to display the data and determine a suitable range and interval size.

Ex. 1: Consider the following list of daily high temperatures in July for Waterloo, ON.

July	1	2	3	4	5	6	7	8	9	10
Temp (°C)	27	25	24	30	32	31	29	24	22	19

July	11	12	13	14	15	16	17	18	19	20
Temp (°C)	21	25	26	31	33	33	30	29	27	28

July	21	22	23	24	25	26	27	28	29	30	31
Temp (°C)	26	27	22	18	20	25	26	29	32	31	28

- a) Is temperature a discrete or continuous variable? Explain.

- b) Should this data be presented in a *grouped* (i.e. intervals) or *ungrouped* (i.e. classes) format?

- c) Determine a range and suitable interval size.

d) Complete the following frequency table.

Temperature Intervals	Midpoint	Tally	Frequency	Cumulative Frequency	Relative Frequency

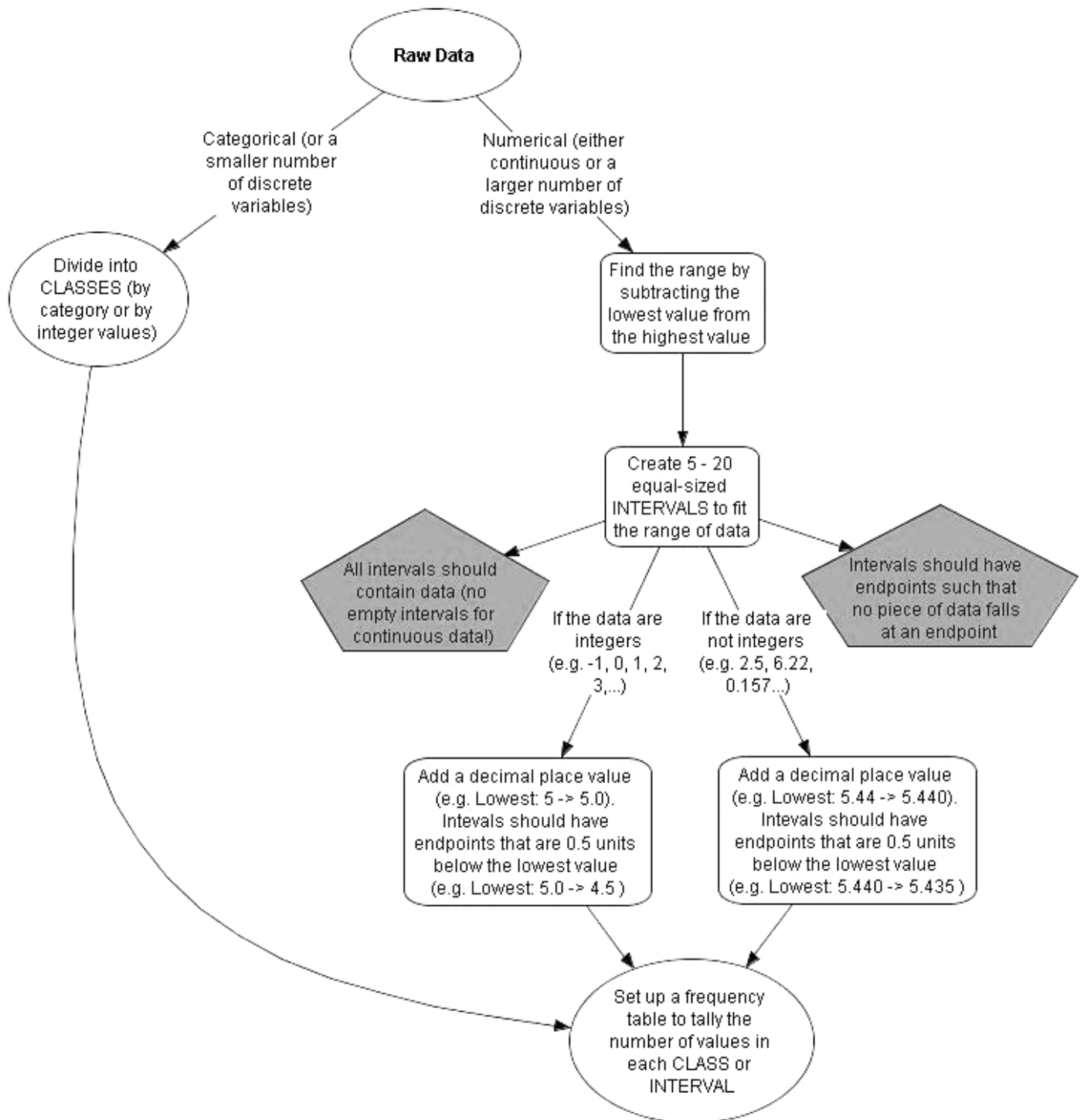
e) Using Excel, recreate the frequency table above, but populate the cells using appropriate formulas. Then construct a frequency histogram, a relative frequency histogram and a cumulative frequency polygon.

f) Use the graphs to answer the following questions:

i. On how many days was the maximum temperature 25°C or less?

ii. What proportion of days was the temperature greater than or equal to 30°C ?

Creating a Frequency Table Using Raw Data



Practise

A

1. Explain the problem with the intervals in each of the following tables.

a)

Age (years)	Frequency
28–32	6
33–38	8
38–42	11
42–48	9
48–52	4

b)

Score (%)	Frequency
61–65	5
66–70	11
71–75	7
76–80	4
91–95	1

2. Would you choose a histogram or a bar graph with separated bars for the data listed below? Explain your choices.

- the numbers from 100 rolls of a standard die
- the distances 40 athletes throw a shot-put
- the ages of all players in a junior lacrosse league
- the heights of all players in a junior lacrosse league

3. A catering service conducted a survey asking respondents to choose from six different hot meals.

Meal Chosen	Number
Chicken cordon bleu	16
New York steak	20
Pasta primavera (vegetarian)	9
Lamb chop	12
Grilled salmon	10
Mushroom stir-fry with almonds (vegetarian)	5

- Create a circle graph to illustrate these data.

- Use the circle graph to determine what percent of the people surveyed chose vegetarian dishes.
 - Sketch a pictograph for the data.
 - Use the pictograph to determine whether more than half of the respondents chose red-meat dishes.
4. a) Estimate the number of hours you spent each weekday on each of the following activities: eating, sleeping, attending class, homework, a job, household chores, recreation, other.
- Present this information using a circle graph.
 - Present the information using a pictograph.

Apply, Solve, Communicate

5. The examination scores for a biology class are shown below.

68	77	91	66	52	58	79	94	81
60	73	57	44	58	71	78	80	54
87	43	61	90	41	76	55	75	49

- Determine the range for these data.
- Determine a reasonable interval size and number of intervals.
- Produce a frequency table for the grouped data.
- Produce a histogram and frequency polygon for the grouped data.
- Produce a relative-frequency polygon for the data.
- Produce a cumulative-frequency polygon for the data.
- What do the frequency polygon, the relative-frequency polygon, and the cumulative-frequency polygon each illustrate best?

B

6. a) Sketch a bar graph to show the results you would expect if you were to roll a standard die 30 times.
 b) Perform the experiment or simulate it with software or the random-number generator of a graphing calculator. Record the results in a table.
 c) Produce a bar graph for the data you collected.
 d) Compare the bar graphs from a) and c). Account for any discrepancies you observe.
7. **Application** In order to set a reasonable price for a “bottomless” cup of coffee, a restaurant owner recorded the number of cups each customer ordered on a typical afternoon.

2	1	2	3	0	1	1	1	2	2
1	3	1	4	2	0	1	2	3	1

- a) Would you present these data in a grouped or ungrouped format? Explain your choice.
 b) Create a frequency table and diagram.
 c) Create a cumulative-frequency diagram.
 d) How can the restaurant owner use this information to set a price for a cup of coffee? What additional information would be helpful?
8. **Application** The list below shows the value of purchases, in dollars, by 30 customers at a clothing store.

55.40	48.26	28.31	14.12	88.90	34.45
51.02	71.87	105.12	10.19	74.44	29.05
43.56	90.66	23.00	60.52	43.17	28.49
67.03	16.18	76.05	45.68	22.76	36.73
39.92	112.48	81.21	56.73	47.19	34.45

- a) Would you present these data in a grouped or ungrouped format? Explain your choice.

- b) Create a frequency table and diagram.
 c) Create a cumulative-frequency diagram.
 d) How might the store owner use this information in planning sales promotions?

9. The speeds of 24 motorists ticketed for exceeding a 60-km/h limit are listed below.

75	72	66	80	75	70	71	82
69	70	72	78	90	75	76	80
75	96	91	77	76	84	74	79

- a) Construct a frequency-distribution table for these data.
 b) Construct a histogram and frequency polygon.
 c) Construct a cumulative-frequency diagram.
 d) How many of the motorists exceeded the speed limit by 15 km/h or less?
 e) How many exceeded the speed limit by over 20 km/h?

10. **Communication** This table summarizes the salaries for François’ hockey team.



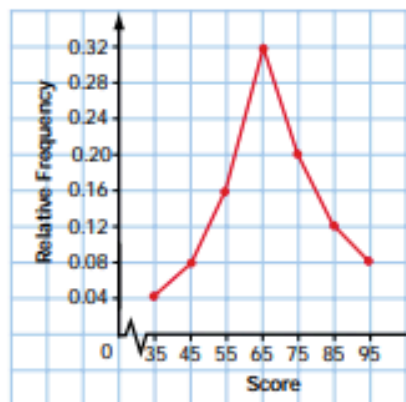
Salary (\$)	Number of Players
300 000	2
500 000	3
750 000	8
900 000	6
1 000 000	2
1 500 000	1
3 000 000	1
4 000 000	1

- a) Reorganize these data into appropriate intervals and present them in a frequency table.
 b) Create a histogram for these data.
 c) Identify and explain any unusual features about this distribution.

11. **Communication**

- What is the sum of all the relative frequencies for any set of data?
- Explain why this sum occurs.

12. The following relative-frequency polygon was constructed for the examination scores for a class of 25 students. Construct the frequency-distribution table for the students' scores.



13. **Inquiry/Problem Solving** The manager of a rock band suspects that MP3 web sites have reduced sales of the band's CDs. A survey of fans last year showed that at least 50% had purchased two or more of the band's CDs. A recent survey of 40 fans found they had purchased the following numbers of the band's CDs.

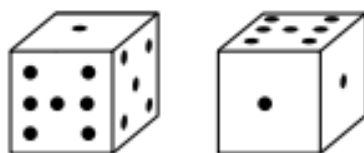
2	1	2	1	3	1	4	1	0	1
0	2	4	1	0	5	2	3	4	1
2	1	1	1	3	1	0	5	4	2
3	1	1	0	2	2	0	0	1	3

Does the new data support the manager's theory? Show the calculations you made to reach your conclusion, and illustrate the results with a diagram.

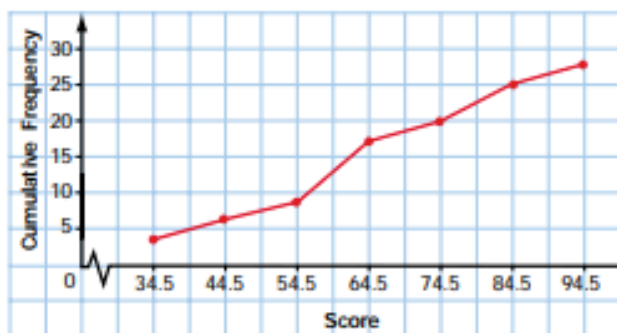


14. **Inquiry/Problem Solving**

- What are the possible outcomes for a roll of two "funny dice" that have faces with the numbers 1, 1, 3, 5, 6, and 7?



- Sketch a relative-frequency polygon to show the results you would expect if these dice were rolled 100 times.
 - Explain why your graph has the shape it does.
 - Use software or a graphing calculator to simulate rolling the funny dice 100 times, and draw a relative-frequency polygon for the results.
 - Account for any differences between the diagrams in parts b) and d).
15. This cumulative-frequency diagram shows the distribution of the examination scores for a statistics class.



- What interval contains the greatest number of scores? Explain how you can tell.
 - How many scores fall within this interval?
16. Predict the shape of the relative-frequency diagram for the examination scores of a first-year university calculus class. Explain why you chose the shape you did. Assume that students enrolled in a wide range of programs take this course. State any other assumptions that you need to make.

Data Analysis with Graphs II

A. Practice Using Excel

Working with Excel, complete the following two questions. Answer all parts and print the answers for your notes. This lesson will give you practice working with data and its organization in Excel.

Ex. 1: For a science project, a student tested how long heavy-duty batteries would power an MP3 player. The student tested 16 different sample batteries and recorded the running time of the MP3 player in hours. The running times are listed here:

29, 26, 23, 22, 22, 17, 27, 25, 22, 19, 23, 22, 27, 23, 24, 26

- Determine the range for these data.
- Determine a reasonable interval size and the number of intervals.
- Produce a frequency table for these data.
- Complete a frequency histogram, a relative-frequency histogram and a cumulative-frequency polygon.

Ex.2: Here is a set of wind speed measurements (in kilometres per hour) recorded hourly on a relatively calm weekend. Prepare a frequency table and appropriate histograms for these data. Remember to calculate the range and select an appropriate interval size.

6	8	10	12	4
4	5	7	7	8
10	5	6	5	6
8	8	8	11	9
7	8	4	4	4
8	10	8	3	4
5	11	9	6	8
10	11	10	5	7
7	8	7	6	6
9	9	7	9	10

Sampling Techniques

A. Gathering and Obtaining Data

Now that we have been introduced to simple graphical tools that will help us summarize data, we will consider the planning of simple sample surveys. A large part of statistics is gathering information; mathematicians and statisticians not only analyse data, but also collect it. Statistics Canada is the body within the Canadian Government responsible for this collection.

A **survey** is used to estimate attributes of human populations. A **census** is a survey that examines *every unit* of the population, such as the Canadian Census Survey conducted every 5 years. Although conducting a census is a very accurate method of acquiring data, it is very costly and time-consuming.

Is there a way we can get highly accurate results without conducting a census?

B. Sampling Terminology

- **Population**

- **Sampling Frame**

- **Sample**

Ex. 1: For your Data Management culminating activity, you wish to look at the effect that having a part-time job has on Ontario 12th grader's school grades. Your principal will not let you survey students outside of your school, and your print account is almost empty (you only have enough to print out 30 surveys!).

Identify the **population**, **sampling frame** and **sample** for your project.

Population:

Sampling Frame:

Sample:

C. Sampling Techniques/Protocols

Once you have identified the population, you need to decide how you will obtain your data. If the population is small, you might be able to complete a census and survey the entire group. For larger populations, it will probably be best to use an appropriate **sampling technique**. If selected carefully, a relatively small sample can give quite accurate results!

There are a number of sampling techniques that we will investigate in this lesson. They include simple random sampling, cluster sampling, convenience sampling, stratified sampling and systematic sampling.

i) Simple Random Sampling (SRS)

Technique	Description	Example
Simple Random Sampling		

Ex. 2: An MDM 4UI student is interested in studying the proportion of students in this class that take public transportation to get to school. Since the student is very busy and does not want to ask *everyone* in the class, the student will conduct the survey only with those students who roll a _____, _____, or _____ on their die.

a) Did the student use SRS?

b) What is the population, sampling frame and sample?

Population:

Sampling Frame:

Sample:

c) Determine the sample proportion of students who use public transportation to get to school. Compare this to the true proportion.

CW: Complete a table comparing the seven sampling techniques outlined on p. 114-116 in your textbook. Practice questions on p. 117 #1-4, 8.

Communicate Your Understanding

1. What are the advantages and disadvantages of using a sample to estimate the characteristics of a population?
2. Discuss whether a systematic sample is a random sample.
3. a) Explain the difference between stratified sampling and cluster sampling.
b) Suggest a situation in which it would be appropriate to use each of these two sampling techniques.

Practise


A

1. Identify the population for each of the following questions.
 - a) Who should be the next president of the students' council?
 - b) Who should be next year's grade-10 representative on the student council?
 - c) What is your favourite soft drink?
 - d) Which Beatles song was the best?
 - e) How effective is a new headache remedy?
2. Classify the sampling method used in each of the following scenarios.
 - a) A radio-show host invites listeners to call in with their views on banning smoking in restaurants.
 - b) The Heritage Ministry selects a sample of recent immigrants such that the proportions from each country of origin are the same as for all immigrants last year.
 - c) A reporter stops people on a downtown street to ask what they think of the city's lakefront.
 - d) A school guidance counsellor arranges interviews with every fifth student on the alphabetized attendance roster.
 - e) A statistician conducting a survey randomly selects 20 cities from across Canada, then 5 neighbourhoods from each of the cities, and then 3 households from each of the neighbourhoods.
 - f) The province randomly chooses 25 public schools to participate in a new fundraising initiative.
3. What type(s) of sample would be appropriate for
 - a) a survey of engineers, technicians, and managers employed by a company?
 - b) determining the most popular pizza topping?
 - c) measuring customer satisfaction for a department store?

Apply, Solve, Communicate

B

4. Natasha is organizing the annual family picnic and wants to arrange a menu that will appeal to children, teens, and adults. She estimates that she has enough time to survey about a dozen people. How should Natasha design a stratified sample if she expects 13 children, 8 teens, and 16 adults to attend the picnic?

5. **Communication** Find out, or estimate, how many students attend your school. Describe how you would design a systematic sample of these students. Assume that you can survey about 20 students.
6. The newly elected Chancellor of the Galactic Federation is interested in the opinions of all citizens regarding economic conditions in the galaxy. Unfortunately, she does not have the resources to visit every populated planet or to send delegates to them. Describe how the Chancellor might organize a multi-stage sample to carry out her survey.
7. **Communication** A community centre chooses 15 of its members at random and asks them to have each member of their families complete a short questionnaire.
- What type of sample is the community centre using?
 - Are the 15 community-centre members a random sample of the community? Explain.
 - To what extent are the family members randomly chosen?
8. **Application** A students' council is conducting a poll of students as they enter the cafeteria.
- What sampling method is the student council using?
 - Discuss whether this method is appropriate for surveying students' opinions on
 - the new mural in the cafeteria
 - the location for the graduation prom
 - Would another sampling technique be better for either of the surveys in part b)?
9. **Application** The host of a call-in program invites listeners to comment on a recent trade by the Toronto Maple Leafs. One caller criticizes the host, stating that the sampling technique is not random. The host replies: "So what? It doesn't matter!"
- What sampling technique is the call-in show using?
 - Is the caller's statement correct? Explain.
 - Is the host's response mathematically correct? Why or why not?
-  C
10. Look in newspapers and periodicals or on the Internet for an article about a study involving a systematic, stratified, cluster, or multi-stage sample. Comment on the suitability of the sampling technique and the validity of the study. Present your answer in the form of a brief report. Include any suggestions you have for improving the study.
11. **Inquiry/Problem Solving** Design a data-gathering method that uses a combination of convenience and systematic sampling techniques.
12. **Inquiry/Problem Solving** Pick a professional sport that has championship playoffs each year.
- Design a multi-stage sample to gather your schoolmates' opinions on which team is likely to win the next championship.
 - Describe how you would carry out your study and illustrate your findings.
 - Research the media to find what the professional commentators are predicting. Do you think these opinions would be more valid than the results of your survey? Why or why not?

Bias in Surveys



Statistical Bias is any factor that favours certain outcomes on responses, hence systematically skewing the survey results. Often bias is unintentional, but as we've seen from the comics, some people deliberately bias surveys in order to get the results they want.

1. Sampling Bias:

This type of bias is why we need to carefully choose a sampling technique!

Ex. 1: Identify the bias in each of the following surveys and suggest how it could be avoided.

- a) A survey asked students in a high school auto-shop whether they own their own car or not.
- b) Students handed out surveys in the school lobby Friday afternoon, asking questions about skipping school.

2. Non-Response Bias:

To avoid non-response bias, researchers often include questions that identify members of particular groups to verify that they are properly represented in the sample. (e.g. gender)

Ex. 2: A science class asks every fifth student entering the cafeteria to answer a survey on environmental issues. Less than half agree to complete the questionnaire. The completed questionnaires show that a high proportion of the respondents are concerned about the environment and are well informed about environmental issues. How is this sample biased?

3. Measurement Bias:

Ex. 3: A highway engineer suggests that an economical way to survey traffic speeds on an expressway would be to have police officers patrol the highway and record the speed of the traffic around them every half hour. How is this sample biased?

4. Response Bias:

Ex. 4: Common instances of response bias...

- a. Blood Donor Clinics

- b. University Professor Evaluations

- c. Teachers asking questions in class

5. Wording:

a. Leading Question:

b. Loaded Question:

Communicate Your Understanding

1. Explain the difference between a measurement bias and a sampling bias.
2. Explain how a researcher could inadvertently bias a study.
3. Describe how each of the following might use intentional bias
 - a) the media
 - b) a marketing department
 - c) a lobby group

Practise

A

1. Classify the bias in each of the following scenarios.
 - a) Members of a golf and country club are polled regarding the construction of a highway interchange on part of their golf course.
 - b) A group of city councillors are asked whether they have ever taken part in an illegal protest.
 - c) A random poll asks the following question: "The proposed casino will produce a number of jobs and economic activity in and around your city, and it will also generate revenue for the provincial government. Are you in favour of this forward-thinking initiative?"
 - d) A survey uses a cluster sample of Toronto residents to determine public opinion on whether the provincial government should increase funding for the public transit.

Apply, Solve, Communicate

2. For each scenario in question 1, suggest how the survey process could be changed to eliminate bias.

3. **Communication** Reword each of the following questions to eliminate the measurement bias.

- a) In light of the current government's weak policies, do you think that it is time for a refreshing change at the next federal election?
- b) Do you plan to support the current government at the next federal election, in order that they can continue to implement their effective policies?
- c) Is first-year calculus as brutal as they say?
- d) Which of the following is your favourite male movie star?
 - i) Al Pacino
 - ii) Keanu Reeves
 - iii) Robert DeNiro
 - iv) Jack Nicholson
 - v) Antonio Banderas
 - vi) Other:
- e) Do you think that fighting should be eliminated from professional hockey so that skilled players can restore the high standards of the game?

B

4. **Communication**
 - a) Write your own example of a leading question and a loaded question.
 - b) Write an unbiased version for each of these two questions.

**ACHIEVEMENT CHECK**Knowledge/
UnderstandingThinking/Inquiry/
Problem Solving

Communication

Application

5. A school principal wants to survey data-management students to determine whether having computer Internet access at home improves their success in this course.
- What type of sample would you suggest? Why? Describe a technique for choosing the sample.
 - The following questions were drafted for the survey questionnaire. Identify any bias in the questions and suggest a rewording to eliminate the bias.
 - Can your family afford high-speed Internet access?
 - Answer the question that follows your mark in data management.
Over 80%: How many hours per week do you spend on the Internet at home?
60–80%: Would home Internet access improve your mark in data management?
Below 60%: Would increased Internet access at school improve your mark in data management?
 - Suppose the goal is to convince the school board that every data-management student needs daily access to computers and the Internet in the classroom. How might you alter your sampling technique to help achieve the desired results in this survey? Would these results still be statistically valid?
6. **Application** A talk-show host conducts an on-air survey about re-instituting capital punishment in Canada. Six out of ten callers voice their support for capital punishment. The next day, the host claims that 60% of Canadians are in favour of capital punishment. Is this claim statistically valid? Explain your reasoning.

C

- Locate an article from a newspaper, periodical, or Internet site that involves a study that contains bias.
 - Briefly describe the study and its findings.
 - Describe the nature of the bias inherent in the study.
 - How has this bias affected the results of the study?
 - Suggest how the study could have eliminated the bias.
8. **Inquiry/Problem Solving** Do you think that the members of Parliament are a representative sample of the population? Why or why not?