MATHEMATICS 9

DATA ANALYSIS & **PROBABILITY**

Section 9.1 – Probability in Society

Probability refers to the likelihood that an event will occur.

By collecting data, predictions can be made about the likelihood that a certain event will occur.

Experimental Probability vs. Theoretical Probability

Experimental Probability is found by repeating an experiment and observing the outcomes.

$$P(\text{event}) = \frac{\text{number of times event occurs}}{\text{total number of trials}}$$

Example:
A coin is tossed 10 times:
A head is recorded 7 times and a tail 3 times.

$$P(\text{head}) = \frac{7}{10}$$

$$P(\text{tail}) = \frac{3}{10}$$

Theoretical Probability is what is expected to happen based on mathematics

$$P(\text{event}) = \frac{\text{number of favorable outcomes}}{\text{total number of possible outcomes}}$$

Example:
A coin is tossed.

$$P(\text{head}) = \frac{1}{2}$$

$$P(\text{tail}) = \frac{1}{2}$$

In general, the higher number of trials you do, the closer the experimental probability will be to the theoretical probability.

How to *express* probability

- Probability can be written in 3 ways...

- As a ***fraction*** = $1/6$
- As a ***decimal*** = $0.1\overline{6}$
- As a ***percent***

  $0.16 \times 100\% = 16\%$

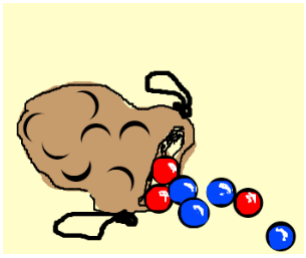
How often will the number 2 show up when rolled?

Example:

Kathleen made 44 out of 50 basketball shots attempted. What is the probability she will sink the next shot? Write the answer as a fraction, decimal and percent.

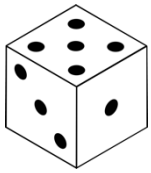
Example:

A bag contains 20 marbles. There are 15 red marbles and 5 blue marbles. What is the theoretical probability that you will pick red?



Example:

a) What is the theoretical probability of rolling a 5 on a die?



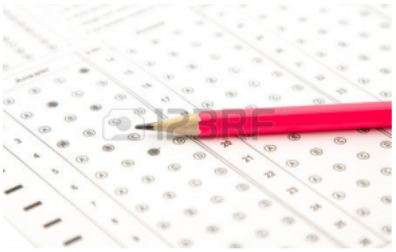
b) In an experiment, you rolled a die 25 times and rolled a 5, five times. What was your experimental probability?

We can use probability to help us make decisions.

Sometimes the decisions we make are influenced by our state of mind or by our gut feeling. When we make a decision in this way, we use **subjective judgement**.

"Buying a lottery ticket today because you feel lucky" , would be an example of using subjective judgment. This means the probaility is based on emotional views and a strong personal opinion. It is like basing probaility on a hunch or a feeling.

Another example of using subjective judgement would be thinking, *"I did well on a my test today because I wrote in pink pencil and pink is my favourite color."*



Example:

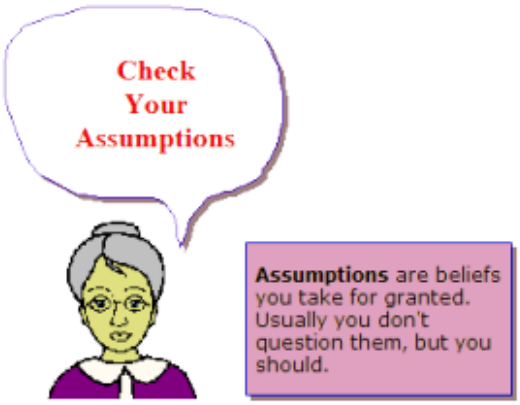
Explain whether each decision is based on theoretical probability, experimental probability or subjective judgement.

a) Listening to a song you dislike on the radio because you feel the next song must be one of your favourites.

b) It is Kate's experience that 4 out of 5 times, the prize in the cereal box is found at the bottom of the box. She opens the cereal from the bottom.

c) Michael is rolling a die 8 times. So far, he has rolled a "4", 7 times in a row. He predicts the next roll will likely not be a 4 since each number has an equal chance of being rolled.

Sometimes when we make a decsion, we take certain things for granted. When we do this, we make **assumptions**.



Example:

A health magazine claims that people who exercise 3 times per week for an hour each time reduce their risk of heart disease by 40%. Stephanie plays sccoer 3 times a week for an hour. So, Stephanie thinks she has reduced her risk of heart disease by 40%.

a) What assumptions is Stephanie making?

b) How might the outcome change if the assumptions change?

Section 9.2 – Potential Problems with Collecting Data

A questionnaire should contain questions that are **fair**. A question should not influence a person's answer. If it does, the question is a **biased question**.

Each person should be able to answer the question.

A biased question could lead to a problem with data collection. There are also many other factors that should be considered.

Potential Problems

- **Bias** – when a question influences answers for or against a topic.

  "Don't you think movie tickets are too high?" – there is already a bias against the price in the question.

- **Use of Language** – the way a question is worded could lead people to answer in a certain way.

  "Don't you think the cafeteria food is too expensive?" – this question may lead students to answer "yes".

  A better wording might be ...

  Do you think food in the caferteria is priced
  a) too low? *b) just right?* *c) too high?*

- **Timing** – when data is collected could affect the results.

  The results of a survey on ski-jackets may differ if the survey is conducted in the summer instead of the winter.

- **Privacy** – people may not want to share personal information.

  Students may not want share their school grades when asked.

- **Cultural Sensitivity** – the survey may offend other cultures.

  A question about Christmas shopping may offend those who do not celebrate Christmas.

- **Ethics** – collected data can only be used for the purposes told to the participants, otherwise your actions are considered unthical.

  Surveying your class to find out their favourite snacks so you can have them available at your birthday party. But then you use the information to sell your classmates snacks between classes

- **Cost** – the cost of collecting data must be considered.

  The cost of mailing a survey to a large number of people may be too expensive.

- **Time** – the amount of time needed to collect the data should be considered.

  A 30 minute survey conducted over the lunch hour may not interest students. Most would not want to give up 30 minutes of their lunch!

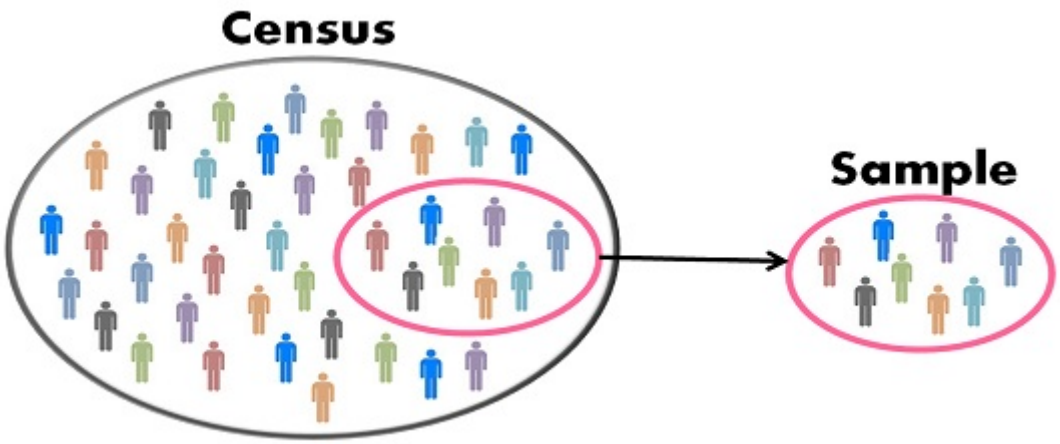
Section 9.3 – Using Samples and Populations to Collect Data

When collecting data, the **population** is the entire group from which you are getting information.

A **census** is conducted when data is collected from **each member** of the population.

A census can be costly, time consuming and difficult or impossible to complete. As a result, a census is only used when an issue is important or when a population is small.

If a census is not feasible or necessary then data is collected from a small portion of the population, which is called a **sample**.



When the sample chosen is representative of the population, that data collection provides **valid conclusions**.

Care must be taken when determining the appropriate size of the sample.

- If the sample is large, the data collection could be costly or time consuming.

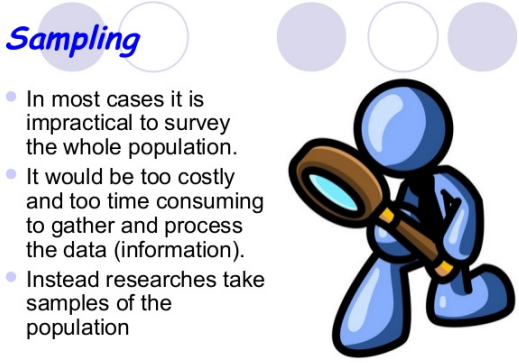
- If the sample is small, then it may not be representative of the population.

Is it always reasonable to survey the entire population?

NO!

Why not?

- Cost
- Time
- Difficult
- Almost impossible to complete



When a population is too large, it is often necessary to collect a representative portion of the population – called a sample.

For example, surveying 50 out of 300 grade 9 students in a large junior high school is a sample.

- **The sample should represent the population**.

For example, the sample should contain a proportional number of boys and girls.

If the grade 9 population of 300 students has 150 boys and 150 girls, then the sample of 50 students should have 25 boys and 25 girls.

- **The size of the sample should also be considered.**

For example, if the sample size is too small the results may not represent the population. 10 grade 9 students may not provide valid conclusions or represent what is typical of the population.

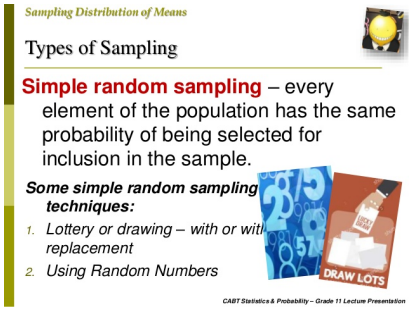
Section 9.4 – Selecting a Sample

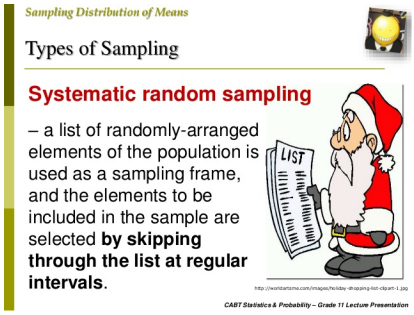
When we cannot survey an entire population, we choose a sample from the population.

Here are some common sampling methods:

- **Simple Random Sampling**
  - Each member of the population has an equal chance of being selected.

  For example, drawing names from a hat.



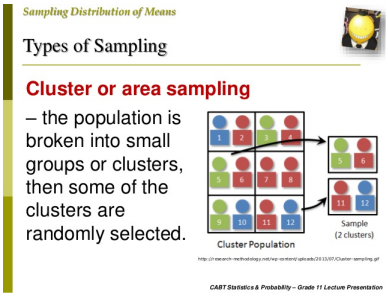


- **Systematic or Interval Sampling**
  - Every n^{th} member of the population is selected.

  For example, choosing every 10^{th} person.

- **Cluster Sampling**
  - Every member of each randomly chosen group of the population is selected.

  For example, one grade is chosen randomly and then all students in that grade are selected.



Self-selected sample

select only members of the population who volunteer for the sample

- **Self-Selected Sampling**
  - Only members who are interested and volunteer will participate.

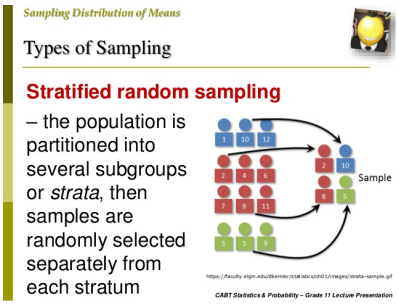
For example, for a radio station telephone survey only people interested will call in.

- **Convenience Sampling**
  - Only members of the population who are convenient to include are selected.

For example, a teacher chooses the first 5 students she sees.

Convenience Sample

select any members of the population who are conveniently and readily available



- **Stratified Random Sampling**
  - Some members from each group of the population are randomly selected.

For example, choosing 5 random students from each grade.

Section 9.5 – Designing a Project Plan

To design a plan to collect data you have to choose a topic that interests you then follow these 5 steps.

1. **Prepare a question**

   - The wording should not influence a person's answer.
   - The question should be sensitive to different cultures.
   - Participants should be anonymous if the question is personal.

2. **Identify the population and possibly choose a sample**

- Decide if you will use a sample or a census.
- Choose a sampling method (consider time and cost).
- Make sure the sample is representative of the population.

3. **Collect the data**

- Make sure the timing of the survey is appropriate.

4. **Analyze and display the data**

- Display the data in a table and/or an appropriate graph.
- Make conclusions from your data.

5. **Evaluate your plan**

- Have you avoided all potential problems with data collection?
- Was your sample appropriate?
- Are your conclusions valid?