Data Analysis Unit

Curriculum Outcomes:

Statistics and Probability (Data Analysis): Collect, display, and analyze data to solve problems.

1. Describe the effect of:

• bias

• use of language

• ethics

• cost

• time and timing

• privacy

• cultural sensitivity

on the collection of data

3. Develop and implement a project plan for the collection, display and analysis of data by:

• formulating a question for investigation

• choosing a data collection method that includes social considerations

• selecting a population or a sample

• collecting the data

• displaying the collected data in an appropriate manner

• drawing conclusions to answer the question.

2. Select and defend the choice of using either a population or a sample of a populatio to answer a question

Things You Need to Know

- **Data is often collected through taking a survey or a questionnaire.**
- **There can be problems with collecting data if we aren't careful. The data we collect might not be accurate.**
- **One way data isn't accurate is if we have a flawed survey question.**
- **Survey questions can be flawed in the following ways:**
    - **<u>Bias:</u> the survey question leans the person taking the survey towards a particular response.**
        - **Ex: *Do you prefer gross, sugary, unhealthy soft drinks, or do you prefer delicious and nutritious milk?***
    - **<u>Language:</u> the survey question might be difficult to understand, or misleading.**
        - **Ex: *Do you not suppose that the least great Prime Minister in Canadian history isn't the one we don't not have right now?***
    - **<u>Ethics:</u> the survey might refer to or deal with inappropriate or illegal content.**
        - **Ex: *Suppose a survey promised to give a free digital copy of a song away without getting the rights from the publisher first.***
    - **<u>Cost:</u> is what the survey is looking for actually worth the effort being put into giving the survey?**
        - **Ex: *Suppose a company paid $1 million to conduct a survey by mail to find out whether people think the letter A is better than the letter B.***
    - **<u>Time/Timing:</u> is the time or place of the survey appropriate, or will it alter people's responses?**
        - **Ex: *Conducting a survey regarding your favourite type of meat at a vegetarian convention.***

        - **Ex:** ***Conducting an annual survey about the noise level at your school, but this year the survey is given during construction season.***
    - **<u>Privacy:</u> can participants refuse to do the survey/can they remain anonymous?**
        - **Ex:** ***Conducting a survey asking people how much money they make, without being allowed to remain anonymous.***
    - **<u>Cultural Sensitivity:</u> is your survey question inclusive to everyone, or does it ignore people of certain cultures/religions/lifestyles?**
        - **Ex:** ***"Which is your favourite meat: beef, chicken, or pork?"***
            - ***This doesn't provide an option for people who don't eat meat.***
- **A <u>population</u> is simply all of the people/things that are the target of your survey.**
- **A <u>sample</u> is just a subgroup of your population that actually takes part in the survey.**
- **Sometimes, you want to/are able to survey the entire population.**
    - **Ex:** ***If you'd like to survey what everyone in your group wants for lunch, you'll survey the entire population (i.e. the group).***
- **Usually, we just survey a sample of the population.**
    - **Ex:** ***If you want to find out favourite movie genres in your school, you may only want to ask a sample of people, because if your school is large it would take a long time to sample the school population.***
- **Types of samples:**
    - **<u>Voluntary Response:</u> participants have the choice of whether or not to participate**
    - **<u>Convenience:</u> participants are chosen because of how easy it was to access them**
    - **<u>Random Samples:</u>**

        - **True Random:** participants are chosen completely randomly (usually by names drawn, or computer generated)
        - **Stratified:** population is split into groups, and the same fraction of the group size is surveyed in each group.
        - **Systematic:** every other person/object in the population is surveyed (ex: every 100th light bulb made in a factory is tested)
- **Depending on what you're looking for, you're going to use a different type of sample.**
- **A biased sample is a sample that doesn't reflect the population.**
    - **Usually, biased samples are due to having a sample size that is too small. A good rule of thumb is having a sample of at least 30.**
- **Measures of Central Tendency:**
    - **mean: add all the data up, and divide by the number of data items**
    - **median: order from least to greatest, find the middle number**
        - **If there are two middle numbers, take the mean of those two middle numbers.**
    - **mode: the number/object that appears most often in the list**