**8.1 Collecting Data – Overview**

**Population vs Sample**

Ex) Lets say I want to figure out if Indiana is a cat or dog state.

**Population** – The particular group of interest in a study (the set of all individuals/objects of interest).

* Ex) Every person in ALL of Indiana.

**Sample** – A subset of individuals/objects from the population of interest.

* Ex) Everybody in Muncie ONLY.

Diagram

Description automatically generated

**Parameter vs Statistic**

**Population parameter** – A fixed numerical value that describes the population.

* Ex) Overall percentage who prefer cats for IN
* Would have to take a **census** (ask everyone in the population) to know this value(or estimate it).

**Sample Statistic** – A numerical value that describes the sample that can vary from sample to sample.

* Ex) Percentage for Muncie (will be different than for Indy)

**Sampling Techniques**

**GOAL: Representative Sample** – A sample that has the same relevant characteristics as the population and does not favor one group of the population over another.

Matches / resembles the population

Graphical user interface

Description automatically generated

Chart, scatter chart, box and whisker chart

Description automatically generated

**1) Random Sample** – A random sample is one in which every member of the population has an equal chance of being selected.

* This is generally desirable but can be difficult to achieve.

**2) Stratified Random Sample** – Dividing population into homogeneous (similar characteristics) groups. This guarantees the sample is representative!

1. Stratify the population – Divide the population into similar groups (e.g. based on age or gender).
2. Take random sample from each group (strata).
3. Combine the groups from each strata to form your sample.

**3) Cluster Sample** – Dividing population into mini-populations. This gives us an unbiased sample and is often a more practical / affordable method.

1. Split the population into representative groupscalled clusters (resemble overall population).
2. Use random sampling to select several clusters.
3. Perform a census of each selected (collect data from every member).

**4) Systematic Sample** –Selecting every nth member of the population.

**5) Convenience Sample** – Include individuals who are convenient to sample (for the researcher); AVOID!

* The group may not be representative of the population 🡪 Frequently ends in biased results.

**Examples**: Describe how you could obtain a sample to answer each question below using each of the following types of sampling methods listed below.

**Example 1**

Scenario: I wish to determine the proportion of the MATH 125 class that has a Mac laptop.

* Random Sample:
* Stratified Random Sample:
* Cluster Sample:
* Systematic Sample:
* Convenience Sample:

**Example 2**

Scenario: You are tasked with conducting a survey to answer the question, “What is the favorite subject of students who attend East High School?”

* Random Sample:
* Stratified Random Sample:
* Cluster Sample:
* Systematic Sample:
* Convenience Sample: