**Stats – Chapter 4**

Ways to sample

Some definitions:

* **Population**

The entire group involved in the case study

* **Census**

Survey that aims to count the population for certain characteristics

* **Sample**

Subset of population on which data is collected

This is considered because analyzing the whole population could be costly, time ineffective and inaccurate

**Simple Random Sampling (SRS)**

SRS is a sampling policy in which each possible sample of *n* size has the same probability of being selected

It doesn’t involve any prior knowledge of the population

To perform SRS, we need to:

1. Define the population (case study)
2. Decide *n*, the sample size
3. Randomly select the sample, either via:
   1. Lottery (aka, “drawing from a hat”)
   2. Random numbering (aka, “random number generator”)
4. Collect data

When using a simple rnaodm sample of *n* subjects:



**Cluster Random Sample**

1. Divide the population into a large number of clusters
2. Select a simple random sample of the clusters
3. Use the subjects in those clusters as the sample

**Stratified Random Sample**

1. A stratified random sample divides the population into separate groups, called strata
2. Selects a simple random sample from each stratum

**Sampling bias**

Sampling bias happens when the probability of being selected is non-uniform, resulting in a biased sample that makes the statistic erroneous

This includes lying (response bias) or abstaining (non-response bias – missing data)

Examples of poor sampling are volunteering and convenience sampling, which are **non-random** (there is a certain type of people targeted and looked for)

**Diagram

Description automatically generated**This doesn’t mean they are not used (e.g. medical studies)

Data can be gathered in two different ways:

* **Experiment study – conducting experiments**

Experimental studies are ones where researchers introduce an intervention and study the effects

Experimental studies are usually randomized, meaning the subjects are grouped by chance, remembering that randomization eliminates the effect of **lurking variables** (causes spurious correlation)

* **Observational study – mere observation**

Observational studies are ones where researchers observe the effect of a risk factor, diagnostic test, treatment, or other intervention without trying to change who is or isn’t exposed to it (i.e. without anything being done to the subjects)

Text

Description automatically generatedThere kind of studies are sensitive to lurking variables

Text

Description automatically generatedThis is used in medical studies, when it wouldn’t be ethical to perform an experiment. We can’t randomly assign subjects into a smoking group and a non-smoking group—this would involve asking some subjects to start smoking

In a case-control study, the number of cases and the number of controls is fixed. The random part is observing the outcome for the explanatory variable