**S12**

**Chapter 5: Organization of Data for Analysis**

**Numerical Data = quantitative data** = data in the form of any number.

* Numerical data can be either discrete (whole numbers) or continuous (decimals, fractions).

**Categorical Data = qualitative data =** data that can be sorted into distinct groups of categories

* Categorical data can be either ordinal (data that can be ranked) or nominal (data that cannot be ranked).

**Keep in mind:**

* Depending on who is analyzing the data and their intentions, the information taken from the date can be very different.
* Variability in data exists due to errors in measurement or varying conditions in experiments.
* Different people can interpret data in different ways.
* When researchers collect data on more than one variable, they can compare the data to see if there is a relationship.

**Population =** all individuals or objects in a group that is being studied.

**Sample** = a group of items or individuals selected from the population.

**Variability in samples** – shows how samples are different from each other. The more similar the samples are to each other, the lower the variability and the more accurately the samples represent the population.

**Simple random sample =** every person/item in the population has an equal chance of being selected.

|  |  |
| --- | --- |
| Type of Sample | Example |
| **Simple Random** |  |
| Systematic |  |
| Stratified |  |
| Cluster |  |
| Convenience |  |
| Voluntary |  |

**Collecting Data**

* Observational study
* Experiment
* Measurement
* Survey

**Treatment group =** the participants in an experiment who receive the specific treatment being measured/tested.

**Control group =** the participants in an experiment who do not receive the specific treatment being measured/tested. Control group is compared to the treatment group.

Note: It is important to randomize the members in the treatment and control group. It is essential to have similar demographic make-up in each group.

**Surveys** – well designed surveys are an effective way to obtain information about a population. A survey should be anonymous, clear, concise and free of bias. Rating scales on a survey should be evenly distributed between good and bad outcomes.

**Bias** occurs when there is prejudice for or against an idea or response. Biased samples can result from problems with either sampling technique or the data collection method.

**Primary source of data** = data that have been collected directly by the researcher and have not been manipulated or summarized.

**Microdata =** an individual set of data about a single respondent.

**Secondary source data** = data used by someone other than those who actually collected them.

 **Aggregate data** = data that are combined or summarized in such a way that the individual microdata can no longer be determined.