

Data Collection and Frequency Distributions

- Trouble with Design
- Sampling Types
- Frequency Distributions

Design Issues

- Voluntary Response
- Sample Size
- Wording of Survey Questions
- Question and Response Order

Observational vs. Experimental

- **Observational Study:** Observe and measure without modifying the subjects.
- **Experiment:** Apply a treatment and then observe its effects on the subjects

Random Samples

- **Random Sample:** Any individual is just as likely to be chosen as any other individual.
- **Simple Random Sample:** Every possible group of n individuals is just as likely to be chosen as any other group of n individuals.

Type of Sampling

- **Systematic:** Select every k^{th} individual, such as every 10th person.
- **Convenience:** Select whatever is easiest.
- **Stratified:** Subdivide the population into groups and establish quotas to ensure that each group has the same proportionate representation in the sample as it has in the population.
- **Cluster:** Divide the population into many sectors. Then randomly select a few sectors and choose **all** members from these chosen sectors.

Timed Studies

- **Cross-sectional:** One point in time is chosen to make all observations.
- **Retrospective:** Data are collected by going back in time.
- **Prospective:** Data are collected in the future from groups (cohorts) sharing common factors.

Blind Studies

- **Control Study:** The population is divided into a control group and a treatment group.
- **Blind Study:** The subjects do not know whether they are part of the control or the treatment group.
- **Double Blind Study:** Neither the subjects nor the treatment provider knows who is in which group.

Error

- **Sampling Error:** The difference between the sample statistic and the population parameter.
- **Nonsampling Error:** An error that occurs when the sample is incorrectly collected, recorded, or analyzed.

Frequency Distributions

- **Frequency Distribution:** A table that shows the number of values in each category

Age	Frequency
15-24	19
25-34	8
35-44	5

Frequency Distribution Attributes

- **Lower Class Limits:** The smallest numbers in each class.
- **Upper Class Limits:** The largest numbers in each class.
- **Class Boundaries:** The numbers separating the classes (the midpoint of the upper class boundary and the next lower class boundary).
- **Class Midpoints:** The midpoint of the lower and upper boundary in each class.
- **Class Width:** The difference between two consecutive lower class boundaries.

$$\text{Width} = \frac{\text{Maximum} - \text{Minimum}}{\text{Number of Classes}} \uparrow$$

Relative and Cumulative

- **Relative Frequency Distribution:** Instead of frequency, use
- **Cumulative Frequency Distribution:** Instead of frequency use the number at or below that class.

Normal Distribution

- The data is symmetric.
- The frequencies start small, get larger and then end small.