**Statistics** is the science of collecting, organizing, and interpreting data.

**Statistics** are the data that describe or summarize something.

Population – the complete set of people or things being studied.

 Ex. How many people are watching American Idol in USA?

Sample – the subset of the population from which the raw data are actually obtained.

 Ex. How many people are watching American Idol in 1 specific State?

Population parameters – specific characteristics of the population that a statistical study is designed to estimate.

 Ex. Characteristics of the viewers.

 Ex. Studies of college costs: all college and universities (p), tuitions, fees, and housing (parameters)

Sample statistics – numbers or observations that summarize the raw data.

 Ex. Percentage of young viewers watching American Idol.

**Example 1 - Population and Sample**: Describe the population, sample, population parameters and sample statistics.

1. Agricultural inspectors Jefferson Country measure the levels of residue for three common pesticides on ears of corn from each of the 104 corm-producing farms in the county.

*(p) – All ears of corn in the country.*

*(s) – 25 ears of corn from each farm.*

*(pp) – The average levels of residue from the 3 pesticides on all corn grown in the county.*

 *(ss) – The average levels of residues that are actually measured on the corn in the sample.*

1. Anthropologists determine the average brain size of early Neanderthals in Europe by studying skulls found at three sites in southern Europe. (a species of human in the genus Homo that became extinct between 41,000 and 39,000 years ago).

*(p) – Population of early Neanderthals.*

*(s) – The relatively few individual Neanderthals whose skulls are found at the three sites.*

*(pp) – Average brain size of all Neanderthals.*

*(ss) – The average brain size (skull size) of the individuals in the sample.*

The Basic Steps in a Statistical Study

1. State the specific goal of the study precisely. Population and the exactness you want to learn.
2. Choose a representative sample from the population.
3. Collect raw data from the sample and summarize these data by finding sample statistics of interest.
4. Use the sample of statistics to infer the population parameters.
5. Draw conclusions. Determine what you learned and whether you achieved your goal.

**Common Sampling Methods: What about bias?**

**Simple random sampling:** we choose a sample of these items in such a way that every sample of a given size has an equal chance of being selected.

**Systematic sampling:** We use a simple system to choose sample, such as selecting every 10th or every 50th member in the population.

**Convenience sampling:** We use a sample that is convenient to select, such as people who happen to be in the same classroom.

**Stratified sampling:** We use this method when we are concerned about differences among subgroups, or strata, within a population. We first identify the subgroups and then draw a simple random sample within each subgroup. The total sample consists of all the samples from the individual subgroups.

* No matter how a sample is chosen, the study can be successful only if the sample is representative of the population.
* Even if the sample is chosen in the best possible way, it is still just a sample (as apposed to the entire population). In general, the larger sample is more likely to be representative of the population, as long as it is chosen well.

A **representative sample** is a sample in which the relevant characteristics of the sample members match those of the population.

**Example 2 - Sampling Methods:** Identify the type of sampling used and comment on whether the sample is likely to be representative of the population.

1. You are conducting a survey of students in a dormitory. You choose you sample by knocking n the door of every 10th room.
* *Choosing every 10ths rooms makes this a systematic sampling. The sample method may be representative, as long as students were randomly assigned to rooms.*
1. To survey opinions on a possible property tax increase, a research firm randomly draws the addresses of 150 homeowners from a public list of all homeowners.
* *The records presumably list all homeowners, so drawing randomly from the list produces a simple random sampling. It has a good chance of being representative of the population.*
1. Agricultural inspectors for Jefferson County check the levels of residue from three common pesticides on 25 ears of corn from each of the 104 corn-producing farms in the county.
* *Each farm may have use different pesticide, so the inspectors consider corn from each farm as a subgroup of the full population. By checking 25 ears of corn from each of the 104 farms, the inspectors are using stratified sampling. If the ears of corns are collected randomly on each farm, each set of 25 is likely to be representative of its farm.*
1. Anthropologists determine that average brain size of the early Neanderthals in Europe by studying skulls found at three sites in southern Europe.
* *By studying skulls found at selected sites, the anthropologists are using a convenient sampling. They have little choice because only a few skulls remain from the many Neanderthals who once lived in Europe. However, it seems reasonable to assume that these skulls are representative of larger population.*

**Types of Statistical Study**

**Observational study** – researchers observe or measure characteristics of the sample members but do not attempt to influence or modify these characteristics.

**Experimental study** – researchers apply a treatment to some or all of the sample members and then look to see whether the treatment has any effects.

**Treatment group** – the group of sample members who receive the treatment being tested.

**Control group** – the group of sample members who do not receive the treatment being tested.

\*\* *It is important for the treatment and control groups to be selected randomly and to be alike in all respects except for the treatment.*

A **placebo** lacks the active ingredient of a treatment being tested in a study, but identical in appearance to the treatment. The study participants cannot distinguish the placebo from the real treatment.

**Placebo effect** refers to the situation in which patients improve simply because they believe they are receiving a useful treatment.

**Blinding** is the practice of keeping people in the dark about who is in the treatment group and who is in control group.

**Single-blind** – the participants do not know whether they are members of the treatment or control group, but the experimenters do know. Example?

**Double-blind** – neither the participants nor the experimenters know who belongs to the treatment group and who belongs to the control group. Example?

**Example 3 - What’s wrong with this experiment? Identify the problems and explain how the problem could have been avoided.**

1. A chiropractor wants to know if his adjustments relieve back pain. He performs adjustments on 25 patients with back pain. Afterward, 18 of his patients say they feel better. He concludes that the adjustments are an effective treatment.
* *All 25 patients received adjustments represent a treatment group, but this study lacks a control group. The patients may be feeling better because of a placebo effect rather than any real effect of the adjustment. The study might be improved by hiring an actor to perform fake adjustment on a control group. Then the chiropractor could have compared the results in the 2 groups to see whether a placebo effect was involved.*
1. A new drug for attention deficit disorder (ADD) is supposed to make the affected children more polite. Randomly selected children suffering from ADD are divided into treatment and control groups. The experiment is single-blind. The experimenters interview the children 1-on-1 to decide whether they became more polite.
* *Because the experimenters know which children received the real drug, during the interviews they may inadvertently speak differently or interpret behavior differently with these children. In that case, their conclusions might not be valid. The experiment should have been double blind, so that the experimenters conducting the interviews would not have known which children received the real drug and which children received placebo.*

**Case-Control Studies** is an observational study that resembles an experiment because the sample naturally divides into 2 or more groups. The participants who engage in the behavior under the study form the **cases**, which make them like a treatment group in an experiment. The participants who do not engage in the behavior are the **controls**, making them like a control group in an experiment.

**Example 4 - Which Type of Study?** What type of statistical study is most likely to lead to an answer? Why?

1. What is the average income of stock brokers?
* *An observational study because we only survey/observe the brokers.*
1. Do seat belts save lives?
* *It is unethical to do so an experiment in which some people were told to wear seat belts and others were told not to wear seat belts. Instead, we can conduct observational case-control study by comparing the death rates in accidents between cases (seat belts) and controls (no seat belt). We can learn whether seat belts save lives. (They do.)*
1. Can lifting weights improve runners’ time in a 10-kilometer race?
* *We need an experiment to determine whether lifting weights can improve runners’ 10k time. What are the criteria for experimental study?*
1. Can a new herbal remedy reduce the severity of colds?
* *Double-blind because we have treatment group, control group and interviewers.*

**Margin of error** describes a **confidence interval** that is likely to contain the true population of parameter.

**Confidence interval** is from (sample statistic – margin of error) to (sample statistic + margin of error)

**Example 5 Close Election**

An election eve poll finds that 52% of surveyed voters plan to vote for Smith, and need a majority (more than 50%) to win without a runoff. The margin of error in the poll is 3 percentage points. Will she win?

1. Find the confidence interval

*From (52% - 3%) to (52% + 3%)*

*From 49% to 55% is the confidence interval.*

1. Draw your conclusion

*The confidence interval leaves open to possibility of both a majority and less than a minority, this election is too close to call.*

**Homework: Quick Quiz 1-10**

 **Does It Make Sense? 9, 11, 13, 15, 17, 19, 29, 31**