AP Statistics Chapter 5 – Producing Data

5.1: Designing Samples

Observation versus Experiment

- An **observational study** observes individuals and measures variables of interest but does not attempt to influence the responses.
- An **experiment**, on the other hand, deliberately imposes some treatment on individuals in order to observe their responses.

Population and Sample

The entire group of individuals that we want information about is called the **population**.
A **sample** is a part of the population that we actually examine in order to gather information.

Sampling versus a Census

Sampling involves studying a part in order to gain information about the whole.
A **census** attempts to contact every individual in the entire population.

Voluntary Response Sample

A **voluntary response sample** consists of people who choose themselves by responding to a general appeal. Voluntary response samples are biased because people with strong opinions, especially negative opinions, are most likely to respond.

Convenience Sampling

A **convenience sample** chooses the individuals easiest to reach. This will typically result in a biased sample of like-minded individuals.

Bias

The design of a study is **biased** if it <u>systematically favors certain outcomes</u>.

Simple Random Sample

A **simple random sample** (SRS) of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.

Random Digits

A **table of random digits** is a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these two properties:

1. Each entry in the table is equally likely to be any of the 10 digits 0 through 9.
2. The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

Choosing an SRS

Choose an **SRS** in two steps:

Step 1: Label. Assign a numerical label to every individual in the population.
Step 2: Table. Use Table B to select labels at random.

Stratified Random Sample

To select a **stratified random sample**, first divide the population into groups of similar individuals, called *strata*. Then choose a separate SRS in each stratum and combine these SRS's to form the full sample.

Undercoverage and Nonresponse

- **Undercoverage** occurs when some groups in the population are left out of the process of choosing the sample.
- **Nonresponse** occurs when an individual chosen for the sample can't be contacted or does not cooperate.

Other Forms of Bias

- The behavior of the respondent or of the interviewer can cause **response bias** in sample results. For example, people may respond differently in person than they would on the phone or if the survey was given anonymously.
- The **wording of questions** is the most important influence on the answers given to a sample survey. Confusing or leading questions can introduce strong bias, and even minor changes in wording can change a survey's outcome.

5.2: Designing Experiments

Experimental Units, Subjects, Treatment

The individuals on which the experiment is done are the **experimental units**. When the units are human beings, they are called **subjects**. A specific experimental condition applied to the units is called a treatment.

Factors, Levels

The explanatory variables in an experiment are often called **factors**. Many experiments study the joint effects of several factors. In such an experiment, each treatment is formed by combining a specific value (often called a **level**) of each of the factors.

Principles of Experimental Design

The basic principles of statistical design of experiments are

1. **Control** the effects of lurking variables on the response, most simply by comparing two or more treatments.
2. **Randomize** – use impersonal chance to assign experimental units to treatments.
3. **Replicate** each treatment on many units to reduce chance variation in the results. More data, more accuracy. We get more trustworthy results.

Statistical Significance

An observed effect so large that it would rarely occur by chance is called **statistically significant**.

Completely Randomized Design

When all experimental units are allocated at random among all treatments, the experimental design is **completely randomized**. Lurking variables are not considered in this design.

Matched Pairs Design

- **Matched pairs** designs compare just two treatments. We choose blocks of two units that are as closely matched as possible. In Example 5.16, two boards on the same pole form a block. We assign one of the treatments to each unit by tossing a coin or reading odd and even digits from Table B.
- Alternatively, each block in a **matched pairs** design may consist of just one subject, who gets both treatments one after the other. Each subject serves as his or her own control. This eliminates the lurking variable of individual differences among the subjects. The order of the treatments can also influence the subject's response, so we randomize the order for each subject, again by a coin toss or other similar method.

Block Design

A **block** is a group of experimental units or subjects that are known before the experiment to be similar in some way that is expected to affect the response to the treatments. In a **block design**, the random assignment of units to treatments is carried out separately within each block. The purpose of blocking is to group similar subjects, thus reducing variation in the observed results.

Double-Blind Experiment

In a double-blind experiment, neither the subjects nor the people who have contact with them know which treatment a subject received. This eliminates human bias on both sides of the experiment.