

### 3.3 Sampling Design

**Population:** the entire group of individuals that we want information about.

**Sample:** a part of the population that we actually examine in order to gather information.

**Bias:** systematic error, in favoring some parts of the population over others.

**Voluntary Response Sample:** consists of people who choose themselves by responding to a general appeal. It's biased because people with strong opinions, especially negative opinions, are most likely to respond.

**Sample design:** refers to the method used to choose the sample from the population.

#### 1. Simple random samples (SRS)

A 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.

Two steps to do a simple random sampling:

- Label all the individuals in the population.
- Use Table B to select a sample of the desired size.

**Example 1** An academic department wishes to choose a three-member advisory committee at random from the members of the department. To choose an SRS of size 3 from the 28 faculty listed below, first label the members of the population as shown.

---

00	Abbott	07	Goodwin	14	Pillotte	21	Theobald
01	Cicirelli	08	Haglund	15	Raman	22	Vader
02	Cuellar	09	Hohnson	16	Riemann	23	Wang
03	Dunsmore	10	Keegan	17	Rodriguez	24	Wieczorek
04	Engle	11	Luo	18	Rown	25	Williams
05	Fitzpatrick	12	Martinez	19	Salazar	26	Wilson
06	Garcia	13	Nguyen	20	Stone	27	Wong

---

## 2. Systematic random sample

We use an example to illustrate what a systematic random sample is.

**Example 2** Suppose that we must choose 4 address out of 100.  $100/4=25$ . So we can think of the list as four lists of 25 addresses. Choose 1 of the first 25 at random by using Table B. The sample contains this address and the addresses 25, 50, and 75 places down the list from it. If 13 is chosen, for example, then the systematic random sample consists of the addresses numbered 13, 38, 63, and 88.

- Use Table B to choose a systematic random sample of 5 addresses from a list of 200. Enter the table at line 120.
- Like and SRS, a systematic sample gives all individuals the same chance to be chosen. Explain why this is true, then explain carefully why a systematic sample is nonetheless not an SRS.

### Difference between SRS and systematic random sample:

- A SRS not only gives each individual an equal chance to be chosen but gives every possible sample an equal chance to be chosen.
- On the other hand, a systematic random sample gives each individual, but not each sample, an equal chance.

## 3. Stratified Samples

**Probability sample:** a sample chosen by chance. We must know what samples are possible and what chance, or probability, each possible sample has.

**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 SRSs to form the full sample.

Strata for sampling are similar to blocks in experiment.

Example 3 A club has 30 student members and 10 faculty members. The students are

Abel	Elashoff	Hernandez	Kiefer	Neyman	Santos
Carson	Fisher	Holland	Klotz	O'Brien	Shaw
Chen	Golomb	Huber	Liu	Pearl	Thompson
David	Griswold	Jimenez	Moran	Potter	Utts
Deming	Hein	Jones	Moskowitz	Reinmann	Vlasic

and the faculty members are

Andrews	Fernandez	Kim	Moore	Rabinowitz
Besicovitch	Gupta	Lightman	Phillips	Yang

The club can send 4 students and 2 faculty members to a convention and decides to choose those who will go by random selection. Use stratified random sampling to do it.

#### 4. Multistage samples

We use the Current Population Survey of the United States to illustrate what a multistage sample looks like. The rough design for that survey is:

Step 1: Divide the United States into 2007 geographical areas called Primary Sampling Units, or PSUs. Select a sample of 754 PSUs. This sample includes the 428 PSUs with the largest population and a stratified sample of 328 of the others.

Step 2: Divided each PSU selected into smaller areas called “blocks”. Stratify the blocks using ethnic and other information and take a stratified sample of the blocks in each PSU.

Step 3: Sort the housing units in each block into clusters of four nearby units. Interview the households in a random sample of these clusters.

#### 5. Cautions about sample surveys

Possible sources of bias:

- **Undercoverage:** It occurs when some groups in the population are left out of the process of choosing the sample.
- **Nonresponse:** It occurs when an individual chosen for the sample can not be contacted or does not cooperate.
- **Response bias:** It is caused by the behavior of the respondent or of the interviewer.
- **Poorly worded questions:** Wording of questions is one of the most important parts of a good survey sampling.

## Chapter 4 Probability: The Study of Randomness

### 4.1 Randomness

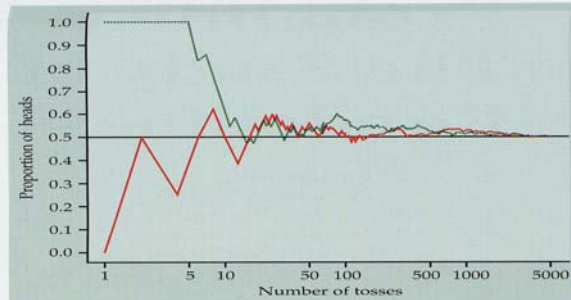


FIGURE 4.1 The proportion of tosses of a coin that give a head changes as we make more tosses. Eventually, however, the proportion approaches 0.5, the probability of a head. Here are the results of two trials of 5000 tosses each.

- **Random:** We call a phenomenon random if individual outcomes are uncertain but there is nonetheless a regular distribution of outcome in a large number of repetitions.
- **Probability:** The probability of any outcome of a random phenomenon is the proportion of times the outcome could occur in a very long series of repetitions. That is, probability is long-term relative frequency.

## 4.2 Probability Models

### 1. Sample space

The sample space  $S$  of a random phenomenon is the set of all possible outcomes.

### 2. Event

1). **Definition:** An event is a subset of the sample space.

- An event expressed in language -----subset from math point of view.
- The relationship between sample space  $S$  and an event  $A$  is the same as the relationship between the set and its subset which can be shown with Venn diagram.

### 2). Operations on events

**Union:** the union of any event  $A$  and event  $B$  is the event that at least one of  $A$  and  $B$  occurs. i.e. the union of  $A$  and  $B = \mathbf{A \text{ or } B}$ , the union of  $A$ ,  $B$  and  $C = \mathbf{A \text{ or } B \text{ or } C}$

**Intersection:** the intersection of any event  $A$  and  $B$  is the event that both  $A$  and  $B$  occur. i.e. the intersection of  $A$  and  $B = \mathbf{A \text{ and } B}$

**Complement:** the complement of any event  $A$  is the event that  $A$  does not occur. i.e. the complement of  $A = \mathbf{A^c}$

### 3). Relationship between events

**Disjoint events:** events A and B are disjoint (or mutually exclusive) if they have no outcomes in common. i.e. A and B never occur simultaneously.

**Independent events:** events A and B are independent if knowing that one occurs does not change the probability that the other occurs.

- Independent events can not be shown by Venn diagram.
- There is no relationship between “disjoint” and “independent”.
- If A and B are independent, then any pair events of  $A^c$  and B, A and  $B^c$ ,  $A^c$  and  $B^c$ , are independent.

### 4). Probability rules

- i)  $0 \leq P(A) \leq 1$  for any event A.
- ii)  $P(S) = 1$
- iii) Complement rule:  $P(A^c) = 1 - P(A)$  for any event A.
- iv) Addition rule for disjoint events: if A and B are disjoint, then  $P(A \text{ or } B) = P(A) + P(B)$ .
- v) Multiplication rule for independent events: if A and B are independent, then  $P(A \text{ and } B) = P(A) \cdot P(B)$ .

Lecture\_7

9

### 5). Assigning probabilities in a finite sample space

Assign a probability to each individual outcome. These probabilities must be numbers between 0 and 1 and must have sum 1.

The probability of any event is the sum of the probabilities of the outcomes making up the event.

▪ **Assigning probabilities by experience.**

▪ **Equally likely outcomes**

- If S contains k outcomes, all equally likely. Then each individual outcome has probability  $1/k$ .
- $P(A) = \text{count of outcomes in } A / \text{count of outcomes in } S$ .

**Example** Screening large numbers of blood samples for HIV, the virus that causes AIDS, uses an enzyme immunoassay (EIA) test that detects antibodies to the virus. Samples that test positive are retested using a more accurate “Western blots” test. Applied to people who have no HIV antibodies, EIA has probability about 0.006 of producing a false positive (that is, of falsely indicating that antibodies are present). If the 140 employees of a medical clinic are tested and all 140 are free of HIV antibodies, what is the probability that at least one false positive will occur?

Lecture\_7

10