Chapter 1

The Nature of Statistics

Section 1.1

Statistics Basics

Definition 1.1

Descriptive Statistics

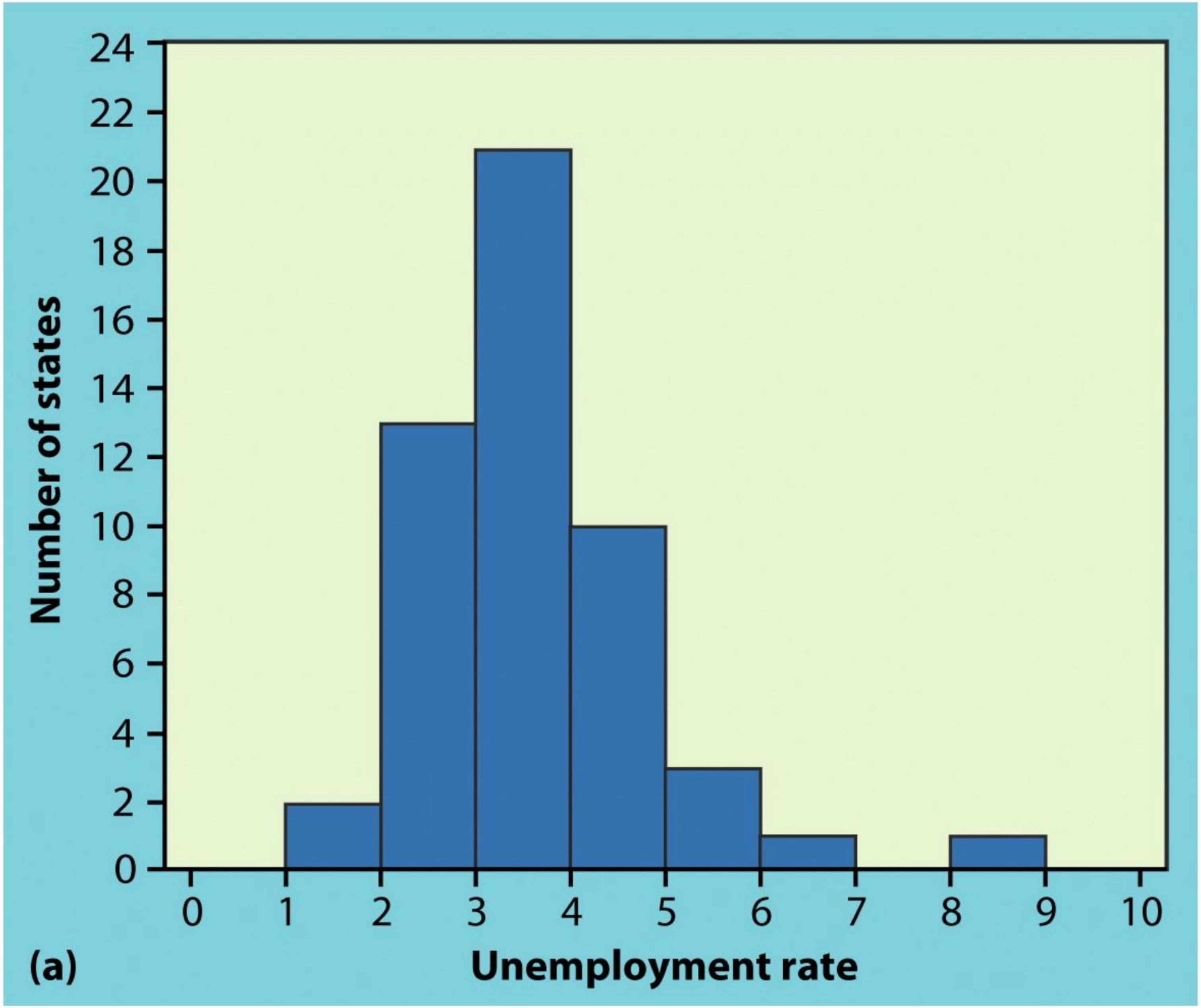
Descriptive Statistics consists of methods for organizing and summarizing information.

Descriptive statistics includes the construction of graphs, charts, and tables and the calculation of various descriptive measures such as averages, measures of variation, and percentiles.

The 1948 Baseball Season. In 1948, the Washington Senators played 153 games, winning 56 and losing 97. They finished seventh in the American League and were led in hitting by Bud Stewart, whose batting average was .279.

TABLE 1.1 Unemployment rates by state, December 2000

State	Percent	State	Percent	State	Percent
Alabama	4.0	Louisiana	5.3	Ohio	3.7
Alaska	6.1	Maine	2.6	Oklahoma	2.6
Arizona	3.3	Maryland	3.3	Oregon	4.0
Arkansas	3.9	Massachusetts	2.0	Pennsylvania	3.8
California	4.3	Michigan	3.4	Puerto Rico	8.9
Colorado	2.1	Minnesota	2.8	Rhode Island	3.2
Connecticut	1.5	Mississippi	4.3	South Carolina	3.3
Delaware	3.3	Missouri	3.2	South Dakota	2.3
Florida	3.2	Montana	4.9	Tennessee	3.8
Georgia	3.0	Nebraska	2.5	Texas	3.4
Hawaii	3.6	Nevada	4.0	Utah	2.7
Idaho	5.0	New Hampshire	2.2	Vermont	2.4
Illinois	4.5	New Jersey	3.5	Virginia	1.9
Indiana	2.7	New Mexico	4.9	Washington	4.9
Iowa	2.5	New York	4.2	West Virginia	5.5
Kansas	3.2	North Carolina	3.6	Wisconsin	3.0
Kentucky	3.7	North Dakota	2.7	Wyoming	3.7



Number of states
24
22
20
18
16
14
12
10
8
6
4
2
0
0
1
2
3
4
5
6
7
8
9
10
(a)
Unemployment rate

24
22
20
18
16
14
12
10
8
6
4
2
0
Number of states
0 1 2 3 4 5 6 7 8 9 10
(b)
Unemployment rate

Definition 1.2

Population and Sample

Population: The collection of all individuals or items under consideration in a statistical study.

Sample: That part of the population from which information is obtained.

Political polling provides an example of **inferential statistics**. Interviewing everyone of voting age in the United States on their voting preferences would be expensive and unrealistic. Statisticians who want to gauge the sentiment of the entire **population** of U.S. voters can afford to interview only a carefully chosen group of a few thousand voters. This group is called a **sample** of the **population**.

Definition 1.3

Inferential Statistics

Inferential statistics: consists of methods for drawing and measuring the reliability of conclusions about a population based on information obtained from a sample of the population.

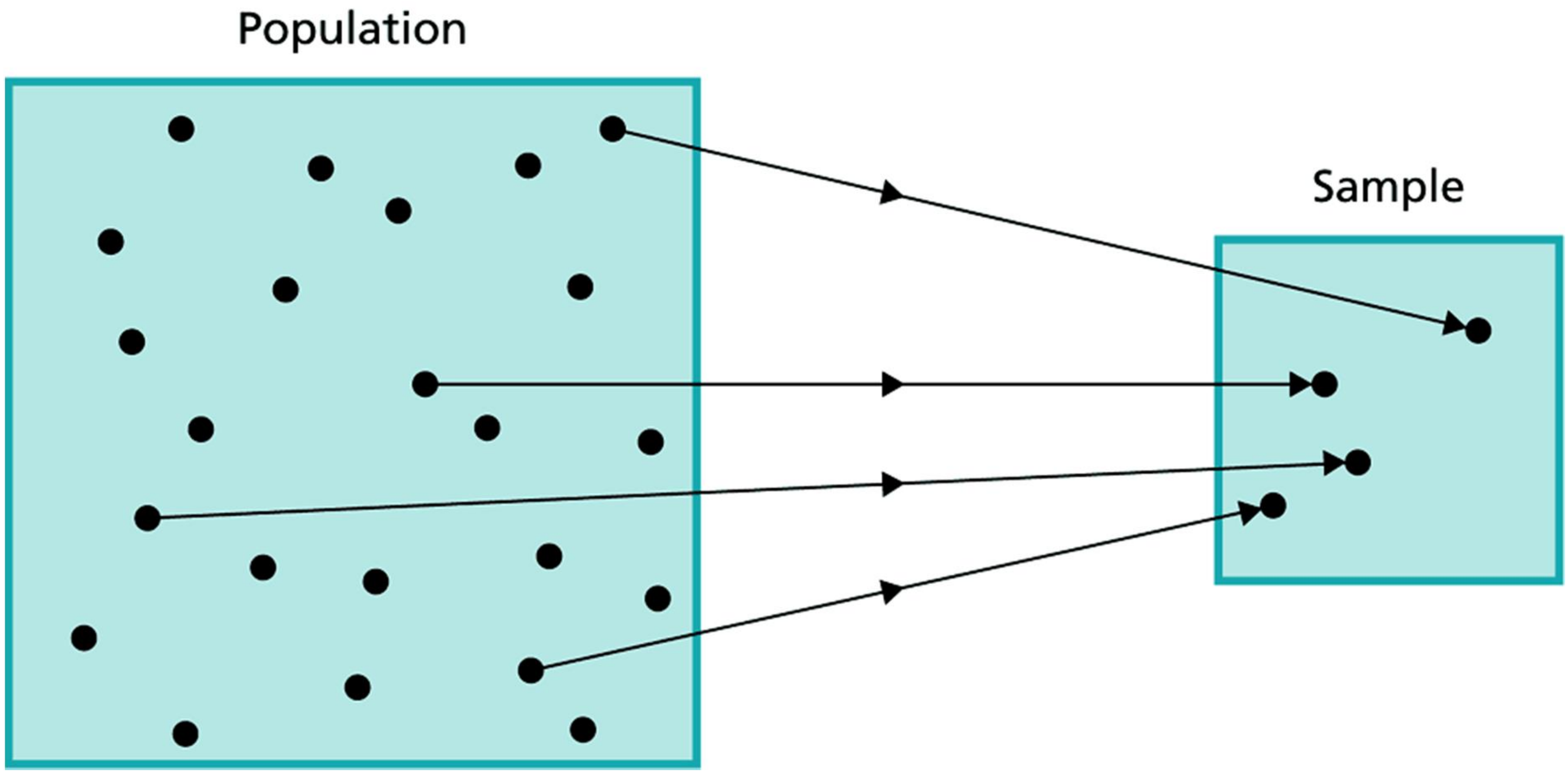
Statisticians analyze the information obtained from a **sample** of the voting **population** to make **inferences** (draw conclusions) about the preferences of the entire voting **population**. Inferential statistics provides methods for drawing such conclusions.

Example of Inferential Statistics

The 1948 Presidential Election In the fall of 1948, President Truman was concerned about statistics. The Gallup Poll taken just prior to the election predicted that he would win only 44.5% of the vote and be defeated by the Republican nominee, Thomas E. Dewey. But the statisticians had predicted incorrectly. Truman won more than 49% of the vote and, with it, the presidency. The Gallup Organization modified some of its procedures and has correctly predicted the winner ever since.

Figure 1.1

Relationship between population and sample



Example 1.3 Classifying Statistical Studies

The 1948 Presidential Election Table 1.1 displays the voting results for the 1948 presidential election

Ticket	Votes	Percentage
Truman–Barkley (Democratic)	24,179,345	49.7
Dewey–Warren (Republican)	21,991,291	45.2
Thurmond–Wright (States Rights)	1,176,125	2.4
Wallace–Taylor (Progressive)	1,157,326	2.4
Thomas–Smith (Socialist)	139,572	0.3

Classification This study is descriptive. It is a summary of the votes cast by U.S. voters in the 1948 presidential election. No inferences are made.

Section 1.2

Simple Random Sampling

Definition 1.4

Simple Random Sampling; Simple Random Sample

Simple random sampling: A sampling procedure for which each possible sample of a given size is equally likely to be the one obtained.

Simple random sample: A sample obtained by simple random sampling.

There are two types of **simple random sampling**.

- One is simple random sampling **with replacement**, whereby a member of the population can be selected more than once;
- the other is simple random sampling **without replacement**, whereby a member of the population can be selected at most once.

Simple Random Sample

- Number each frame unit from 1 to N.
- Use a random number table or a random number generator to select n distinct numbers between 1 and N, inclusively.
- Easier to perform for small populations
- Cumbersome for large populations

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.

Random-Number Tables

Obtaining a simple random sample by picking slips of paper out of a box is usually impractical, especially when the population is large. Fortunately, we can use several practical procedures to get simple random samples. One common method involves a **table of random numbers** – a table of randomly chosen digits, as illustrated in Table 1.5.

Sampling Student Opinions - An Example of Simple Random Samples

A professor want to sample the attitudes of the students taking college algebra at his school.

Method: to interview 15 of the 728 students enrolled in the course. Using a registration list on which the 728 students were numbered 1–728, he obtained a simple random sample of 15 students by randomly selecting 15 numbers between 1 and 728.

Table 1.5

Random numbers

Line number	Column number 00–09		10–19		20–29		30–39		40–49	
00	15544	80712	97742	21500	97081	42451	50623	56071	28882	28739
01	01011	21285	04729	39986	73150	31548	30168	76189	56996	19210
02	47435	53308	40718	29050	74858	64517	93573	51058	68501	42723
03	91312	75137	86274	59834	69844	19853	06917	17413	44474	86530
04	12775	08768	80791	16298	22934	09630	98862	39746	64623	32768
05	31466	43761	94872	92230	52367	13205	38634	55882	77518	36252
06	09300	43847	40881	51243	97810	18903	53914	31688	06220	40422
07	73582	13810	57784	72454	68997	72229	30340	08844	53924	89630
08	11092	81392	58189	22697	41063	09451	09789	00637	06450	85990
09	93322	98567	00116	35605	66790	52965	62877	21740	56476	49296
10	80134	12484	67089	08674	70753	90959	45842	59844	45214	36505
11	97888	31797	95037	84400	76041	96668	75920	68482	56855	97417
12	92612	27082	59459	69380	98654	20407	88151	56263	27126	63797
13	72744	45586	43279	44218	83638	05422	00995	70217	78925	39097
14	96256	70653	45285	26293	78305	80252	03625	40159	68760	84716
15	07851	47452	66742	83331	54701	06573	98169	37499	67756	68301
16	25594	41552	96475	56151	02089	33748	65289	89956	89559	33687
17	65358	15155	59374	80940	03411	94656	69440	47156	77115	99463
18	09402	31008	53424	21928	02198	61201	02457	87214	59750	51330
19	97424	90765	01634	37328	41243	33564	17884	94747	93650	77668

Results:

Random numbers

069 Start
~~988~~
386
539
303
097
628
458
~~759~~
~~881~~
009
036
~~981~~
652
694
024
178
404
578
~~849~~

Not between 1 and 728

Not between 1 and 728

TABLE 1.6
Registration numbers of students interviewed

69	303	458	652	178
386	97	9	694	578
539	628	36	24	404

Random-Number Generators

Nowadays, statisticians prefer statistical software packages or graphing calculators, rather than random-number tables, to obtain simple random samples. The built-in programs for doing so are called **random-number generators**. When using random-number generators, be aware of whether they provide samples with replacement or samples without replacement.

Simple Random Sample: Numbered Population Frame

01 Alaska Airlines	11 DuPont	21 Lucent
02 Alcoa	12 Exxon Mobil	22 Mattel
03 Ashland	13 General Dynamics	23 Mead
04 Bank of America	14 General Electric	24 Microsoft
05 BellSouth	15 General Mills	25 Occidental Petroleum
06 Chevron	16 Halliburton	26 JCPenney
07 Citigroup	17 IBM	27 Procter & Gamble
08 Clorox	18 Kellog	28 Ryder
09 Delta Air Lines	19 KMart	29 Sears
10 Disney	20 Lowe's	30 Time Warner

Simple Random Sampling: Random Number Table

9 9 4 3 7	8 7 9 6 1	4 5 7 3 7	3 7 5 5 2	9 7 9 6 9	3 9 0 9 4	3 4 4 7 5	3 1 6 1 8
5 0 6 5 6	0 0 1 2 7	6 8 3 6 7	6 6 8 8 2	0 8 1 5 6	8 0 0 1 6	7 8 2 2 4	5 8 3 2 6
8 0 8 8 0	6 3 1 7 1	4 2 8 7 7	6 6 8 3 5	6 0 5 1 5	7 0 2 9 6	5 0 0 2 6	4 5 5 8 7
8 6 4 2 0	4 0 8 5 3	5 3 7 9 8	8 9 4 5 4	6 8 1 3 0	9 1 2 5 3	8 8 1 0 4	7 4 3 1 9
6 0 0 9 7	8 6 4 3 6	0 1 8 6 9	4 7 7 5 8	8 9 5 3 5	9 9 4 0 0	4 8 2 6 8	3 0 6 0 6
5 2 5 8 7	7 1 9 6 5	8 5 4 5 3	4 6 8 3 4	0 0 9 9 1	9 9 7 2 9	7 6 9 4 8	1 5 9 4 1
8 9 1 5 5	9 0 5 5 3	9 0 6 8 9	4 8 6 3 7	0 7 9 5 5	4 7 0 6 2	7 1 1 8 2	6 4 4 9 3

- $N = 30$
- $n = 6$

Simple Random Sample: Sample Members

01 Alaska Airlines	11 DuPont	21 Lucent
02 Alcoa	12 Exxon Mobil	22 Mattel
03 Ashland	13 General Dynamics	23 Mead
04 Bank of America	14 General Electric	24 Microsoft
05 BellSouth	15 General Mills	25 Occidental Petroleum
06 Chevron	16 Halliburton	26 JCPenney
07 Citigroup	17 IBM	27 Procter & Gamble
08 Clorox	18 Kellog	28 Ryder
09 Delta Air Lines	19 KMart	29 Sears
10 Disney	20 Lowe's	30 Time Warner

- $N = 30$
- $n = 6$

Section 1.3

Other Sampling Designs

Systematic Sampling

- **Systematic Random Sampling:** The items or individuals of the population are arranged in some order. A random starting point is selected and then every k^{th} member of the population is selected for the sample.

Procedure 1.1

Systematic Random Sampling

Step 1 **Divide the population size by the sample size and round the result down to the nearest whole number, m.**

Step 2 **Use a random-number table or a similar device to obtain a number, k, between 1 and m.**

Step 3 **Select for the sample those members of the population that are numbered $k, k + m, k + 2m, \ldots$.**

Systematic Sampling

- Convenient and relatively easy to administer
- Population elements are an ordered sequence (at least, conceptually).
- The first sample element is selected randomly from the first *k* population elements.
- Thereafter, sample elements are selected at a constant interval, *k*, from the ordered sequence frame.

$$k = \frac{N}{n} ,$$

where:

n = sample size

N = population size

k = size of selection interval

Systematic Sampling: Example

- Purchase orders for the previous fiscal year are serialized 1 to 10,000 ($N = 10{,}000$).
- A sample of fifty ($n = 50$) purchases orders is needed for an audit.
- $k = 10{,}000/50 = 200$
- First sample element randomly selected from the first 200 purchase orders. Assume the 45th purchase order was selected.
- Subsequent sample elements: 245, 445, 645, . . .

Cluster Sampling

- A **cluster sample**: observations in the population are aggregated into larger sampling units, called cluster. Next, an SRS is conducted on the clusters (not the observation), and then sub-sample all or some of the observations in the selected clusters. This process is to make the sampling more convenient or to remedy the fact that we do not have a list of population units, however, this method tends to decrease precision.

Procedure 1.2

Cluster Sampling

Step 1 Divide the population into groups (clusters).

Step 2 Obtain a simple random sample of the clusters.

Step 3 Use all the members of the clusters obtained in Step 2 as the sample.

Cluster Sampling

- Population is divided into nonoverlapping clusters or areas
- Each cluster is a miniature, or microcosm, of the population.
- A subset of the clusters is selected randomly for the sample.
- If the number of elements in the subset of clusters is larger than the desired value of n, these clusters may be subdivided to form a new set of clusters and subjected to a random selection process.

Allocation of the Sample

The best allocation scheme is affected by two factors:

1. The number of clusters in the population.
2. The relative cluster size (i.e. average of M). This value is difficult to estimate, we might need information gained from a prior survey done by a related study or a preliminary sample done by ourselves.

An example of this: An consumer rights inspector wants to estimate the average number of toilette papers per box of a certain brand, packaged at a certain factory. However, the toilette papers is available to the inspector in cartons containing 12 boxes each. Hence the inspector randomly selected x cartons and measures the number of toilette papers for every box in the sampled cartons.

Cluster Sampling

Individual elements of the population are allowed in the sample only if they belong to a cluster. Why use cluster samples?

- Constructing a sampling frame list of observation units may be difficult, expensive, or impossible.
- The population may be widely distributed geographically or may occur in natural cluster such as households or schools.

Stratification – increase precision, units in each stratum are more differ.

Clustering – decrease precision, units in each stratum are more similar. But more cheaper and more convenient.

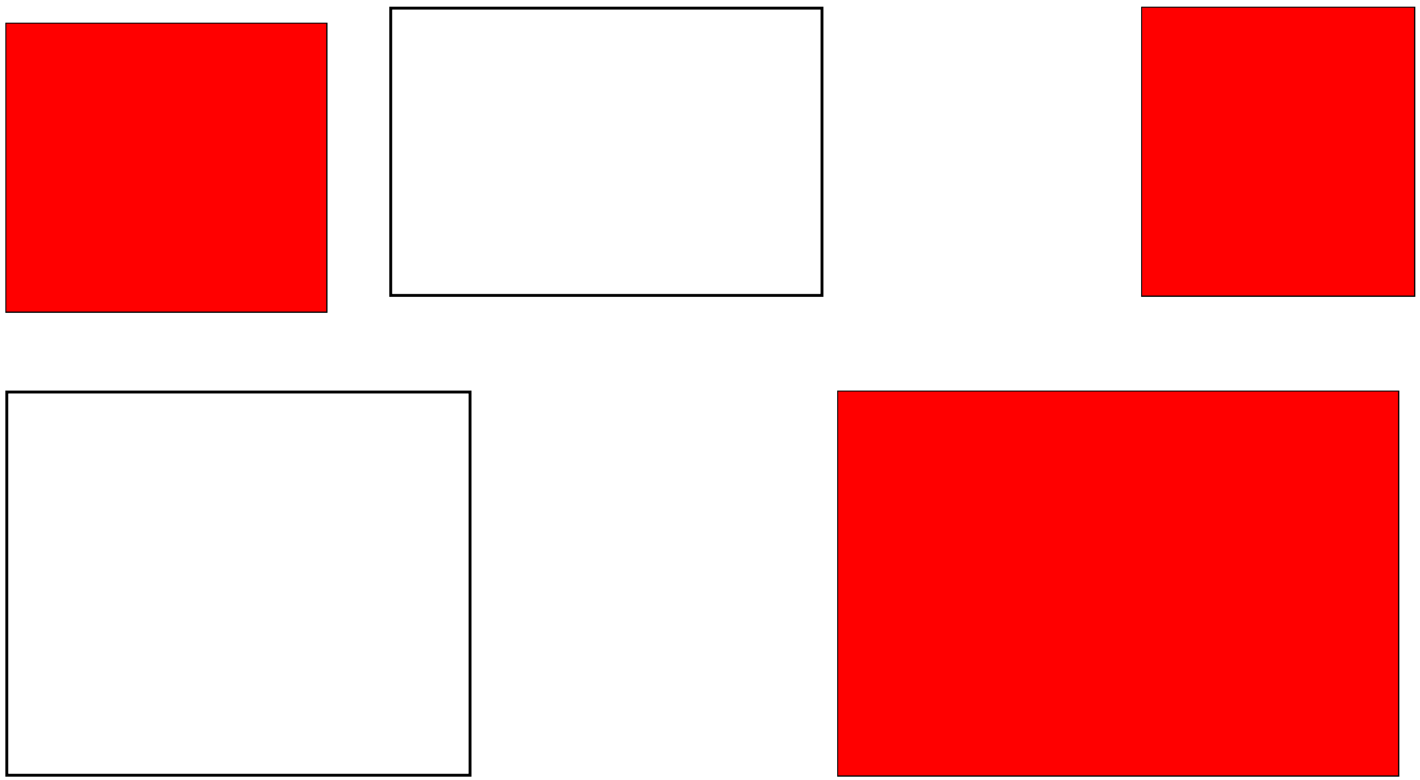
Example 1 on Cluster Sampling

An example that deals with frame: A sociologist wants to estimate the per-capita income of all adult (age 20 or higher) Christians in a city. No list of resident adults is available. How should the sociologist design the sample survey?

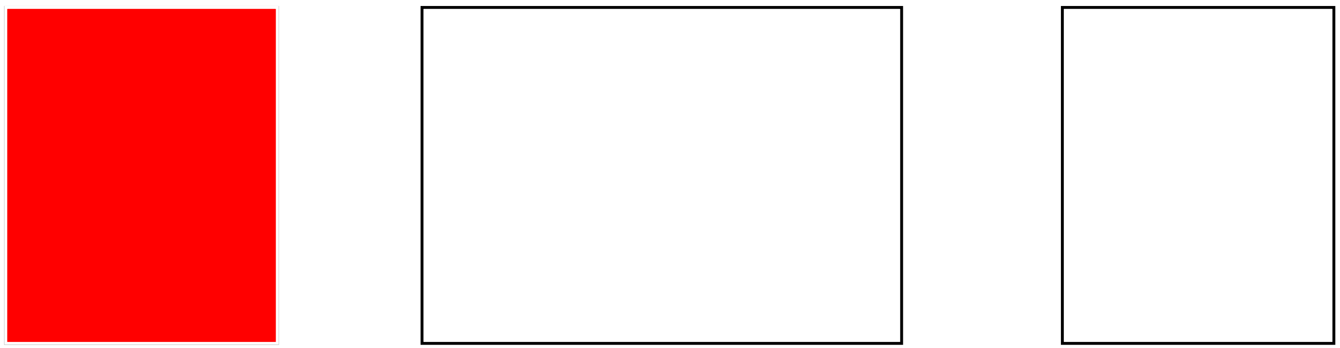
Example 2 on Cluster Sampling

An example that deals with the issue of convenient:
A political scientist developed a test designed to measure the degree of awareness of current events. She wants to estimate the average score that would be achieved on this test by all students in pre-selected universities. Due to administration constraints and technical difficulties, students lists are not available to the study and universities do not want too much disruptions of the classes. How should the scientist design the sample survey?

A diagram to illustrate Cluster Sampling



One Stage Clustering Method



Cluster Sampling

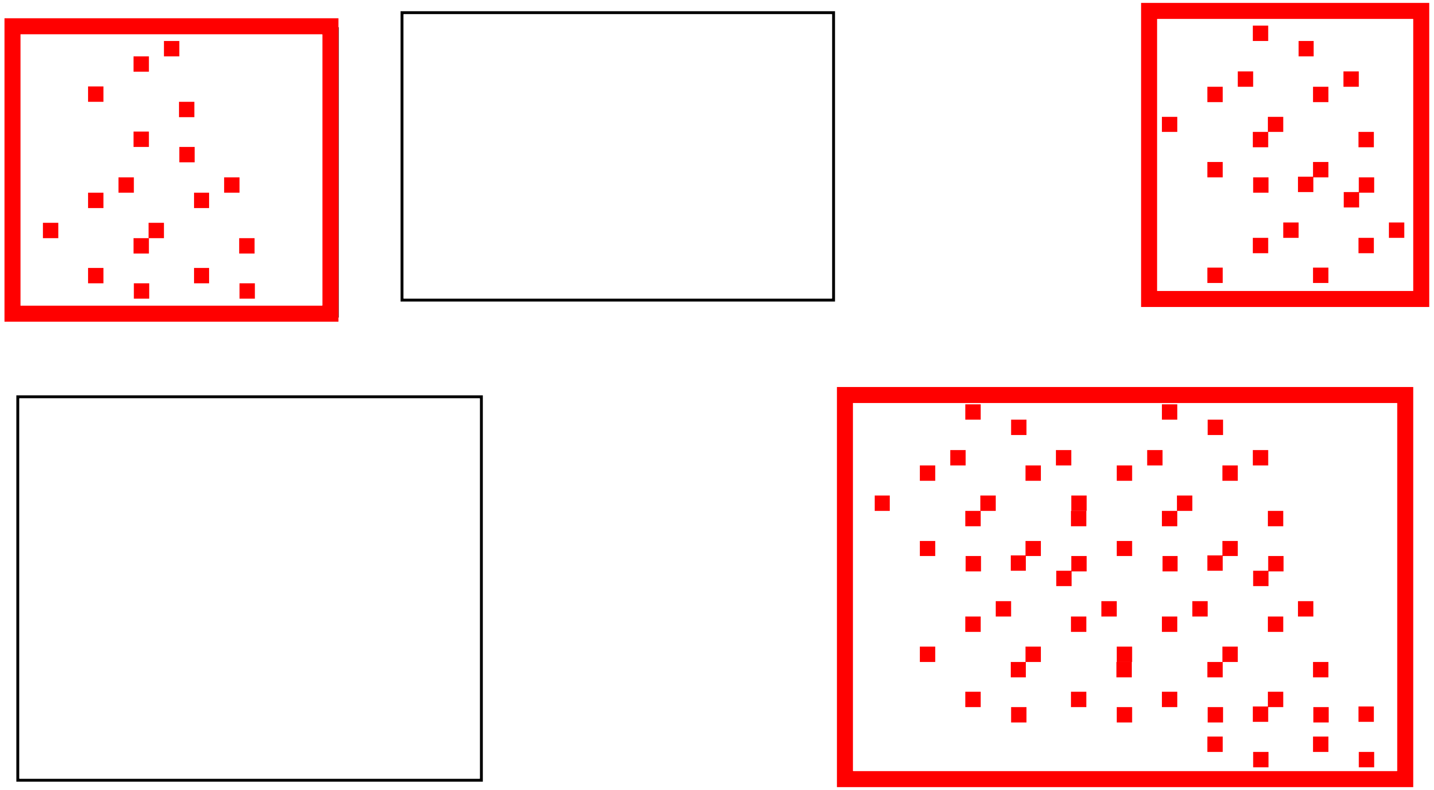
- **Advantages**
  - More convenient for geographically dispersed populations
  - Reduced travel costs to contact sample elements
  - Simplified administration of the survey
  - Unavailability of sampling frame prohibits using other random sampling methods
- **Disadvantages**
  - Statistically less efficient when the cluster elements are similar
  - Costs and problems of statistical analysis are greater than for simple random sampling

Two-stage Cluster Sample

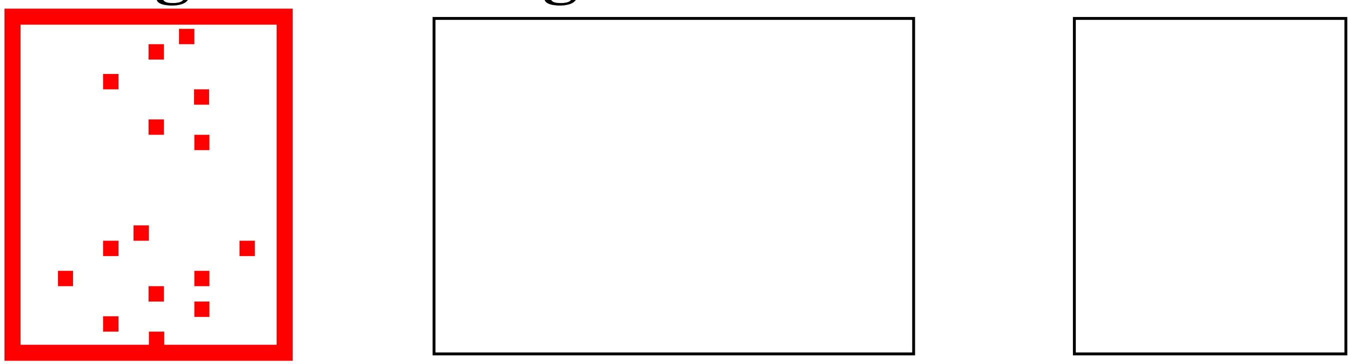
Definition: a two-stage cluster sample is obtained by first selecting a probability sample of clusters and then selecting a probability sample of elements from each sampled cluster.

- Commonly used in large surveys.
- It retains the advantages of cluster sampling.
- It is often more economical.

A diagram to illustrate Cluster Sampling



Two Stage Clustering Method



Stratified Random Sample

- A **stratified random sample**: the population is divided into subgroups, called strata. Then an SRS is selected from each stratum, independently. Usually, elements in the same stratum are more similar, hence this process increases precision.

Procedure 1.3

Stratified Random Sampling with Proportional Allocation

Step 1 Divide the population into subpopulations (strata).

Step 2 From each stratum, obtain a simple random sample of size proportional to the size of the stratum; that is, the sample size for a stratum equals the total sample size times the stratum size divided by the population size.

Step 3 Use all the members obtained in Step 2 as the sample.

Stratified Random Sample

- Population is divided into nonoverlapping groups of similar individuals, called strata
- A random sample is selected from each stratum
- Potential for reducing sampling error
- Proportionate -- the percentage of thee sample taken from each stratum is proportionate to the percentage that each stratum is within the population
- Disproportionate -- proportions of the strata within the sample are different than the proportions of the strata within the population

Stratified Random Sample: Population of FM Radio Listeners

Stratified by Age

20 - 30 years old
(homogeneous within)
(alike)

30 - 40 years old
(homogeneous within)
(alike)

40 - 50 years old
(homogeneous within)
(alike)

Hetergeneous
(different)
between

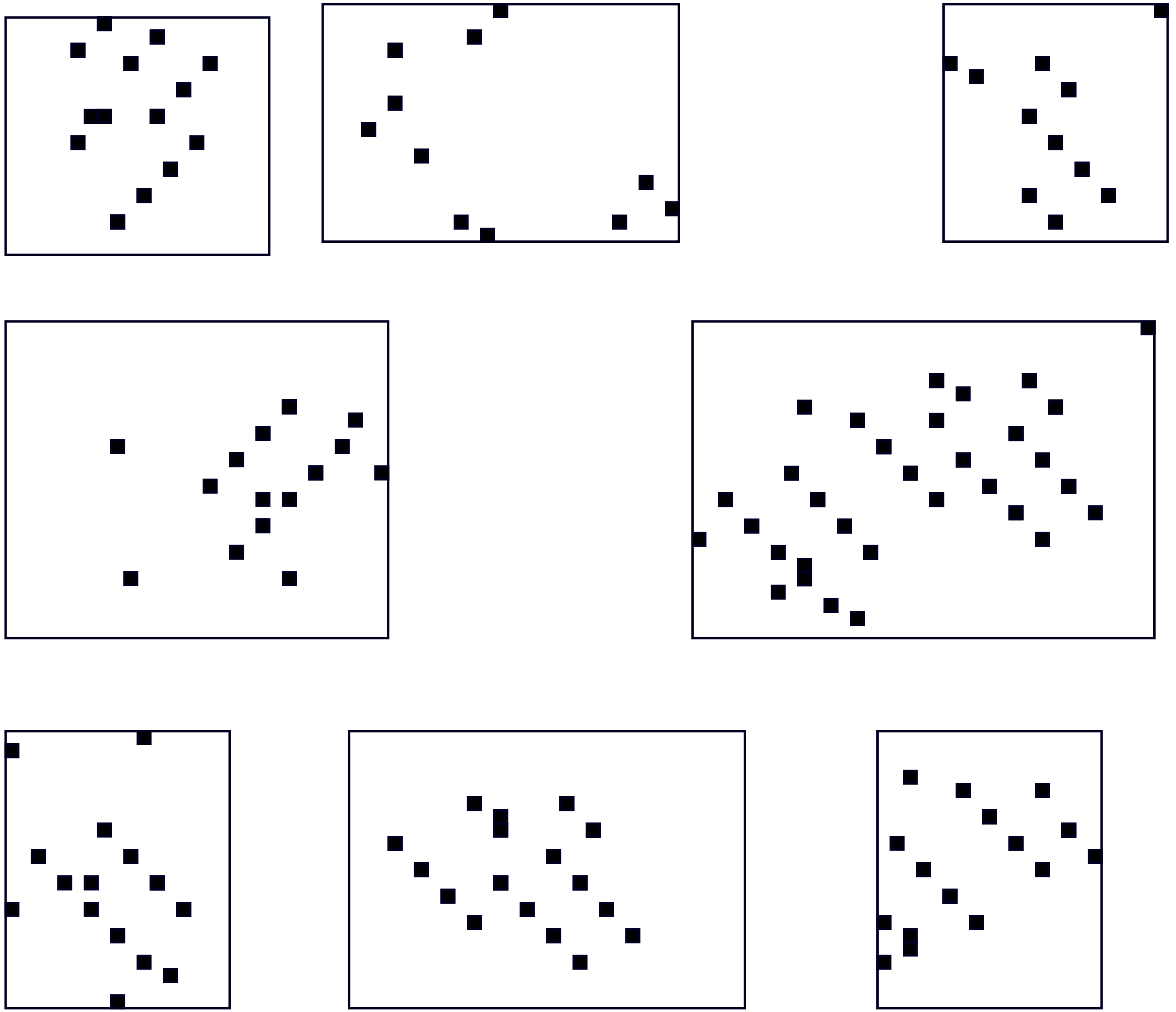
Hetergeneous
(different)
between

Stratified Sampling

Draw independent SRS from each stratum, then pool the information to obtain overall population estimates. The reasons for doing so:

- To be protected from the possibility of obtaining a really bad sample.
- To gain more information for each subgroup.
- It is more convenient to administer and may result in a lower cost for the survey.
- A correctly administered stratified sampling will give more precise estimates for the whole population.

A diagram to illustrate the Stratified Sampling



Allocating Observations to Strata

- Proportional Allocation (self-weighting) – best for similar variances
- Optimal Allocation – strata with higher variance will allocate larger portion of observations.
- Allocation for specified precision within Strata – with respect to given E, σ^2 and expected α.
- Determining Sample Sizes – adjusting after the relative sample size of each strata.
- Post-stratification – stratify after sample has already been taken (with extra information recorded and large sample size).

Stratified sample or SRS?

Answer: added complexity v.s. gain in precision.

Allocation of the Sample

The best allocation scheme is affected by three factors:

1. The total number of elements in each stratum.
2. The variability of the observations within each stratum.
3. The cost of obtaining an observation from each stratum.

Example 1 on Stratified Sampling

An advertising firm finds that obtaining an observation from a rural household, i.e. Taipei county, cost more than obtaining a response from the city, i.e. Taipei city. The increase is due to the cost of traveling from one rural household to another. How would you allocate sample size in this case? Assume the variances for Taipei county and city are similar.

What happen if the advertising firm plans to use telephone interview instead? How would you allocate sample size in this case?

Example 2 on Stratified Sampling

A wholesale food distributor in a large city wants to know whether demand is great enough to justify adding a new product to his stock. To aid in making his decision, he plans to add this product to a sample of the stores he services in order to estimate average monthly sales. Currently, he services four large chain in city. Hence for administrative convenience, he decides to use stratified random sampling with each chain as a stratum. What do you feel about this plan?

Beware of Quota Sampling

Quota samplings are not stratified sampling, since probability sampling is not used to choose simple units in each strata. Choice of sample units is by the discretion of the interviewer or a sample of convenience.

Nonrandom Sampling

- **Convenience Sampling:** *sample elements are selected for the convenience of the researcher*
- **Voluntary Response Sampling:** *survey subjects volunteer to respond to a general appeal. Samples are likely to be biased (such as call-in program, strong opinions are likely to responds).*
- **Judgment Sampling:** *sample elements are selected by the judgment of the researcher.*
- **Quota Sampling:** *sample elements are selected until the quota controls are satisfied.*
- **Snowball Sampling:** *survey subjects are selected based on referral from other survey respondents.*

Section 1.4

Experimental Designs

Definition 1.5

Experimental Units; Subjects

In a designed experiment, the individuals or items on which the experiment is performed are called **experimental units.** When the experimental units are humans, the term **subject** is often used in place of experimental unit.

Folic Acid and Birth Defects. For the study, the doctors enrolled 4753 women prior to conception, and divided them randomly into two groups. One group took daily multivitamins containing 0.8 mg of folic acid, whereas the other group received only trace elements. In the language of experimental design, each woman in the folic acid study is an experimental unit, or a subject.

Key Fact 1.1

Principles of Experimental Design

The following principles of experimental design enable a researcher to conclude that differences in the results of an experiment not reasonably attributable to chance are likely caused by the treatments.

Control: Two or more treatments should be compared.

Randomization: The experimental units should be randomly divided into groups to avoid unintentional selection bias in constituting the groups.

Replication: A sufficient number of experimental units should be used to ensure that randomization creates groups that resemble each other closely and to increase the chances of detecting any differences among the treatments.

Comparative Experiments

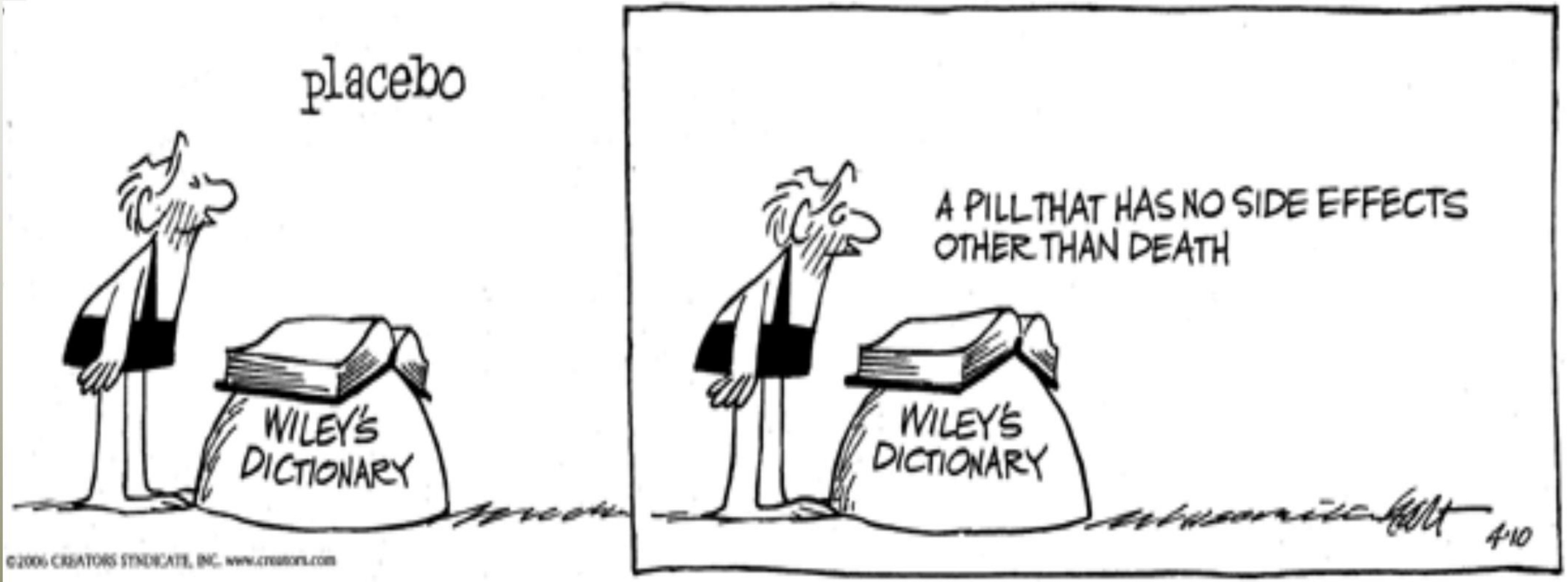
A simple setup:

Subjects -> Treatment -> Response

A poor design - might effect by placebo effect.

Comparative Experiments

Placebo Effect



Comparative Experiments

Placebo effect – responses to a dummy treatment is the placebo effect; might due to the fact that patients trust the doctor and healed by faith.

How do we deal with this effect? Or similar other side-effect in a study?

A better design – use a control group (a sham treatment).

In case of Folic Acid and Birth Defects

Control: The doctors compared the rate of major birth defects for the women who took folic acid to that for the women who took only trace elements.

Randomization: The women were divided randomly into two groups to avoid unintentional selection bias.

Replication: A large number of women were recruited for the study to make it likely that the two groups created by randomization would be similar and also to increase the chances of detecting any effect due to the folic acid.

In case of Folic Acid and Birth Defects

One of the most common experimental situations involves a specified treatment and placebo, an inert or innocuous medical substance. Technically, both the specified treatment and placebo are treatments. The group receiving the specified treatment is called the **treatment group**, and the group receiving placebo is called the **control group**. In the folic acid study, the women who took folic acid constituted the **treatment group** and those who took only trace elements constituted the **control group**.

Definition 1.6

Response Variable, Factors, Levels, and Treatments

Response variable: The characteristic of the experimental outcome that is to be measured or observed.

Factor: A variable whose effect on the response variable is of interest in the experiment.

Levels: The possible values of a factor.

Treatment: Each experimental condition. For one-factor experiments, the treatments are the levels of the single Factor. For multifactor experiments, each treatment is a Combination of levels of the factors.

Example 1.12 Experimental Design: *Weight Gain of Golden Torch Cacti*

The Golden Torch Cactus (Trichocereus spachianus), a cactus native to Argentina, has excellent landscape potential. William Feldman and Frank Crosswhite, two researchers at the Boyce Thompson Southwestern Arboretum, investigated the optimal method for producing these cacti. The researchers examined, among other things, the effects of a hydrophilic polymer and irrigation regime on weight gain. Hydrophilic polymers are used as soil additives to keep moisture in the root zone. For this study, the researchers chose Broadleaf P-4 polyacrylamide, abbreviated P4. The hydrophilic polymer was either used or not used, and five irrigation regimes were employed: none, light, medium, heavy, and very heavy.

Example 1.12 Experimental Design: *Weight Gain of Golden Torch Cacti*

Identify the

a. experimental units.
b. response variable.
c. factors.
d. levels of each factor.
e. treatments.

Example 1.12 Experimental Design: *Weight Gain of Golden Torch Cacti*

Solution

a. The experimental units are the cacti used in the study.
b. The response variable is weight gain.
c. The factors are hydrophilic polymer and irrigation regime.
d. Hydrophilic polymer has two levels: with and without. Irrigation regime has five levels: none, light, medium, heavy, and very heavy.
e. Each treatment is a combination of a level of hydrophilic polymer and a level of irrigation regime. Table 1.8 depicts the 10 treatments for this experiment. In the table, we abbreviated "very heavy" as "Xheavy."

Table 1.8

Schematic for the 10 treatments in the cactus study

Polymer		Irrigation regime				
		None	Light	Medium	Heavy	Xheavy
	No P4	No water No P4 (Treatment 1)	Light water No P4 (Treatment 2)	Medium water No P4 (Treatment 3)	Heavy water No P4 (Treatment 4)	Xheavy water No P4 (Treatment 5)
	With P4	No water With P4 (Treatment 6)	Light water With P4 (Treatment 7)	Medium water With P4 (Treatment 8)	Heavy water With P4 (Treatment 9)	Xheavy water With P4 (Treatment 10)

Definition 1.7

Completely Randomized Design

In a **completely randomized design,** all the experimental units are assigned randomly among all the treatments.

Once we have chosen the treatments, we must decide how the experimental units are to be assigned to the treatments (or vice versa). The women in the folic acid study were randomly divided into two groups; one group received folic acid and the other only trace elements. In the cactus study, 40 cacti were divided randomly into 10 groups of four cacti each and then each group was assigned a different treatment from among the 10 depicted in Table 1.8. Both of these experiments used a **completely randomized design**.

Definition 1.8

Randomized Block Design

In a **randomized block design,** the experimental units are assigned randomly among all the treatments separately within each block.

Although the completely randomized design is commonly used and simple, it is not always the best design. Several alternatives to that design exist. For instance, in a **randomized block design**, experimental units that are similar in ways that are expected to affect the response variable are grouped in **blocks**. Then the random assignment of experimental units to the treatments is made block by block.

Example 1.13 Statistical Designs: *Golf Ball Driving Distances*

Suppose we want to compare the driving distances for five different brands of golf ball. For 40 golfers, discuss a method of comparison based on

a. a completely randomized design.

b. a randomized block design.

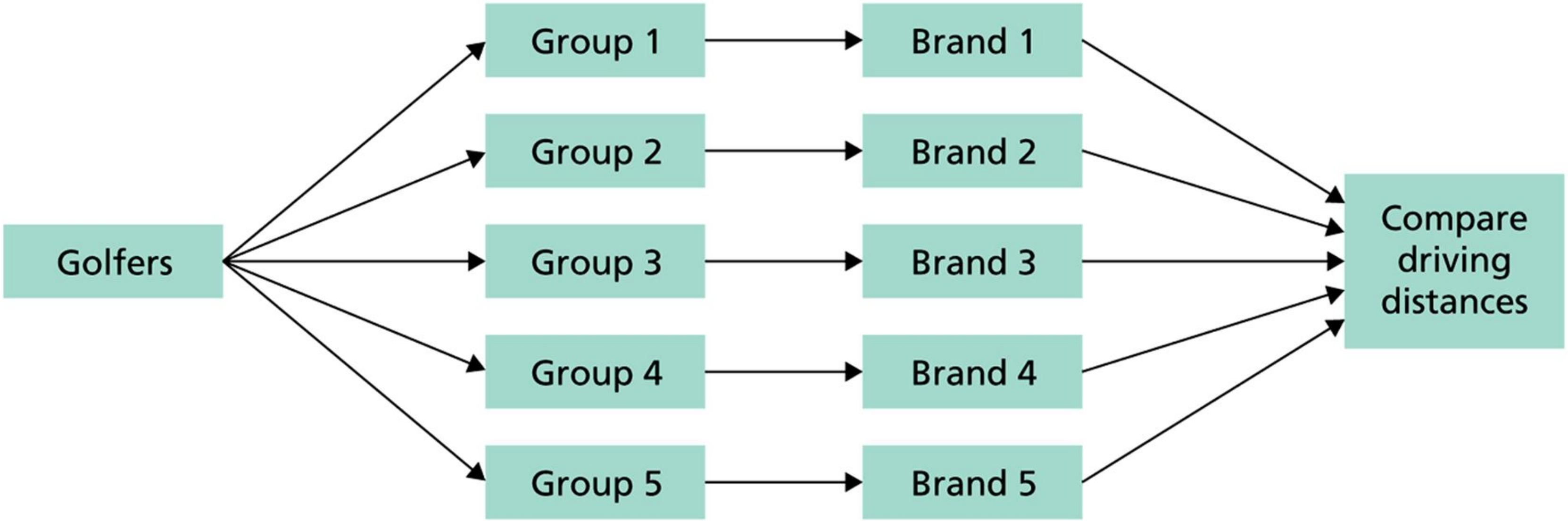
Solution

Here the experimental units are the golfers, the response variable is driving distance, the factor is brand of golf ball, and the levels (and treatments) are the five brands.

a. For a completely randomized design, we would randomly divide the 40 golfers into five groups of 8 golfers each and then randomly assign each group to drive a different brand of ball, as illustrated in Fig.1.5.

Figure 1.5

Completely randomized design for golf ball experiment

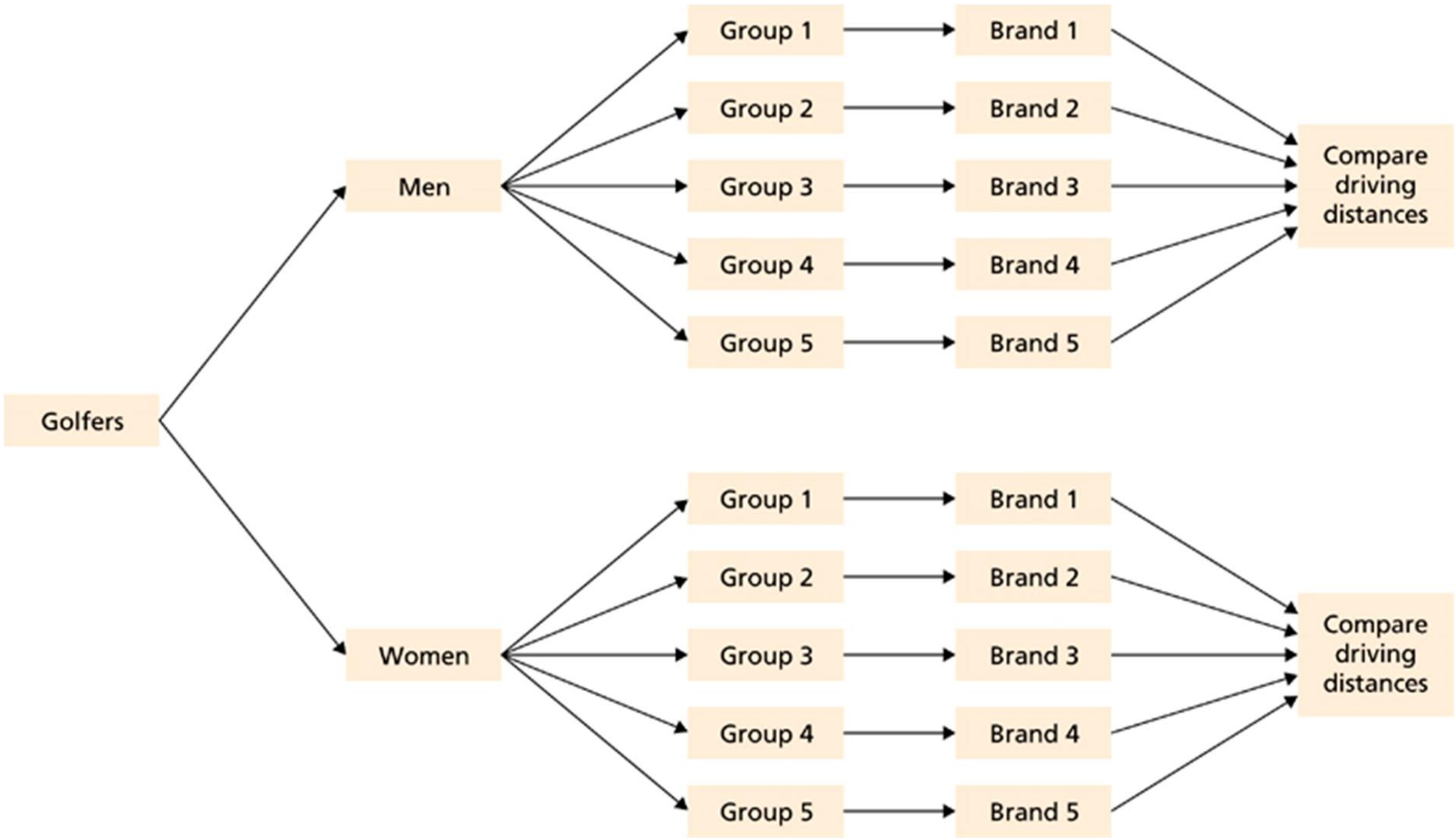


Example 1.13 Statistical Designs: *Golf Ball Driving Distances*

b. Because driving distance is affected by gender, using a randomized block design that blocks by gender is probably a better approach. We could do so by using 20 men golfers and 20 women golfers. We would randomly divide the 20 men into five groups of 4 men each and then randomly assign each group to drive a different brand of ball, as shown in Fig.1.6. Likewise, we would randomly divide the 20 women into five groups of 4 women each and then randomly assign each group to drive a different brand of ball, as also shown in Fig.1.6.

Figure 1.6

Randomized block design for golf ball experiment



By blocking, we can isolate and remove the variation in driving distances between men and women and thereby make it easier to detect any differences in driving distances among the five brands of golf ball. Additionally, blocking permits us to analyze separately the differences in driving distances among the five brands for men and women.

As illustrated in Example 1.13, blocking can isolate and remove systematic differences among blocks, thereby making any differences among treatments easier to detect. Blocking also makes possible the separate analysis of treatment effects on each block.