Section 1.1

Statistics Basics

- The use of statistical methods involves the **collection** of information, **organization**, **analysis** and **interpretation** in order to draw meaningful conclusions

- **Data**: information from observations, counts, measurements or responses

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
- the calculation of various descriptive measures such as averages, measures of variation, and percentiles.
- observing the general nature of the distribution of data

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.

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.

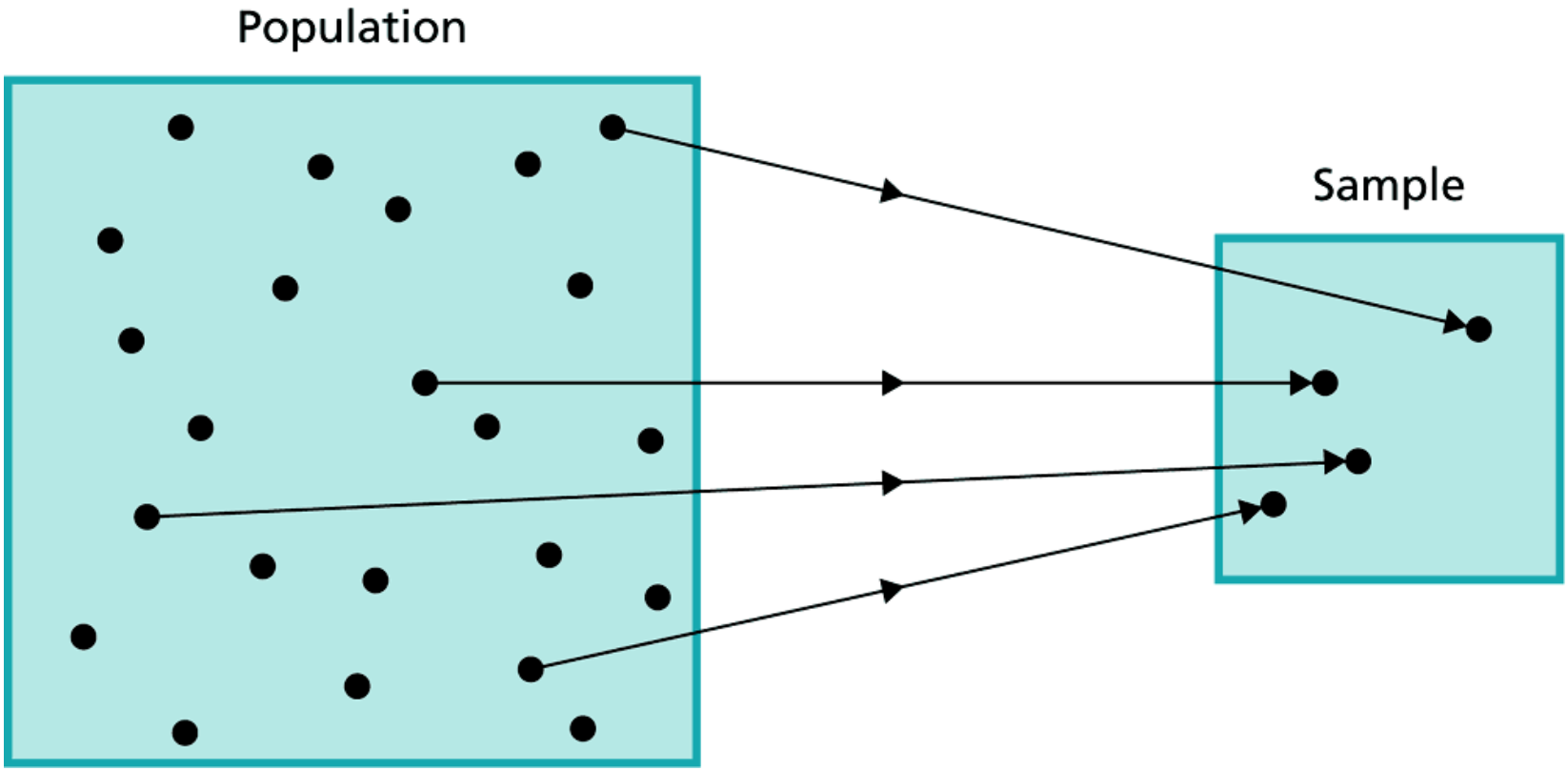
Statisticians analyze the information obtained from a **sample** of the voting **population** to make **inferences** (draw conclusions) about the preferences of the entire voting **population**. Inferential statistics provides methods for drawing such conclusions.

A survey conducted among 1000 people found that 76% of women and 24% of men had a physical exam within the previous year

- Descriptive statistics involve 76% of women and 24% of men who had a physical exam
- Inference drawn from the study is that a higher percentage of women had a physical exam

Figure 1.1

Relationship between population and sample



- **Observational Study**: Observe and measure characteristics of interest – do not change existing conditions

- **Designed Experiment**: Assign treatment and controls (no treatment) groups – observe and compare responses from both groups

Nike developed a new type of sneaker designed to help delay the onset of arthritis in the knee. 80 people with early signs of arthritis volunteered for the study. One-half of them wore the experimental sneaker and the other half wore regular Nike sneakers. Both type of sneakers looked alike. Individuals wore the sneakers every day. At the end of the study, their symptoms were evaluated and MRI tests were performed on their knees. Any differences between those who wore the experimental sneaker and those who did not were evaluated.

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.

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.

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

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.

There are 200 students enrolled in a class.

You want to select 25 students to do a project.

- Assign numbers 001 to 200 to the students.
- Generate 25 random numbers from 1 to 200 using a random number generator or random number tables. Students with those 25 numbers are selected.

Section 1.3

Other Sampling Designs

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 $k, k + m, k + 2m, \ldots$.

- Every 10^{th} person entering a mall is asked to name his/her favorite store

- Every 100^{th} light bulb manufactured is inspected

Example for choosing a systematic sample

Select a sample of size 25 from a population of size 200 using systematic sampling

Step 1 : $m = \frac{\text{Population size}}{\text{Sample size}} = \frac{200}{25} = 8$

Step 2 : Obtain a random # 'K' between 1 and 8
say $K = 2$

Step 3 : Sample members of the population that are numbered $K, K+m, K+2m, K+3m, \ldots$

i.e 2, 10, 18, 26, 34 . . . . . and so for a total of 25 numbers.

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.**

After a hurricane, a disaster area is divided into 200 equal grids. Thirty of the grids (clusters) are randomly selected, and every occupied household in the grid is interviewed.

Example for choosing a cluster sample

From a population of size 50, select a sample of size 20 using cluster sampling assuming the clusters are of equal size 10.

The # of clusters (blocks) each with 10 members of the population is $\frac{50}{10} = 5$

For a sample of size 20, the # of clusters required is $\frac{20}{10} = 2$, each containing 10 members

Randomly select 2 clusters from the 5 existing clusters in the population and get info from everyone in the selected 2 clusters.

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.**

Example of Stratified random sampling using proportional allocation

A town has 250 homeowners of which

- 25 → upper income group
- 175 → middle " "
- 50 → low " "

] three strata

To obtain a sample of 20 homeowners from the three strata using stratified sampling with proportional allocation

$$\#\text{ in sample from upper income group} = n\left(\frac{\#\text{ of high-income homeowners}}{\text{Total \# homeowners}}\right)$$

$$= n\,(\text{proportion of high income homeowners in town})$$

$$= 20\left(\frac{25}{250}\right) = \boxed{2}$$ where 'n' = sample size

$$\#\text{ in sample from middle income group} = 20\left(\frac{175}{250}\right) = \boxed{14}$$

$$\#\text{ in sample from low income group} = 20\left(\frac{50}{250}\right) = \boxed{4}$$

So the sample of size 20 has 2 from upper income, 14 from middle income and 4 from low income homeowners.