Chapter 3: Sampling and Sampling Distributions

Vocabulary:

Population = the set of all the elements of interest in a study

Sample = a subset of the population

Parameters = numerical characteristics of a population such as the mean and SD.

A Main Purpose of Statistical Inference = to develop estimates and test hypotheses about population parameters using information contained in a sample.

(Notes)

****It is important to realize that sample results provide only estimates of the values of the population characteristics.*

Simple Random Sampling

Sampling from a Finite Population

A simple random sample of size *n* from a finite population of size *N* is a sample selected such that each possible sample of size *n* has the same probability of being selected.

(Limited number of population)

Sampling from an Infinite Population

In some situations, the population is either infinite or so large that for practical purposes it must be treated as infinite. For example, suppose that a fast-food restaurant would like to obtain a profile of its customers by selecting a simple random sample of customers and asking each customer to complete a short questionnaire. In such situations, the ongoing process of customer visits to the restaurant can be viewed as coming from an infinite population.

Conditions for Simple Random Sample (Infinite Population)

1. Each element selected comes from the population.
2. Each element is selected independently.

Formula to find the number of different simple random samples: $\dfrac{N!}{n!(N-n)!}$

Table 1: Random Numbers

63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	95436	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15702
46276	87453	44790	67122	45573	84358	21625	16999	13385	22782
55363	07449	34835	15290	76616	67191	12777	21861	68689	03263
69393	92785	49902	58447	42048	30378	87618	26933	40640	16281
13186	29431	88190	04588	38733	81290	89541	70290	40113	08243
17726	28652	56836	78351	47327	18518	92222	55201	27340	10493
36520	64465	05550	30157	82242	29520	69753	72602	23756	54935
81628	36100	39254	56835	37636	02421	98063	89641	64953	99337
84649	48968	75215	75498	49539	74240	03466	49292	36401	45525
63291	11618	12613	75055	43915	26488	41116	64531	56827	30825
70502	53225	03655	05915	37140	57051	48393	91322	25653	06543
06426	24771	59935	49801	11082	66762	94477	02494	88215	27191
20711	55609	29430	70165	45406	78484	31639	52009	18873	96927
41990	70538	77191	25860	55204	73417	83920	69468	74972	38712
72452	36618	76298	26678	89334	33938	95567	29380	75906	91807
37042	40318	57099	10528	09925	89773	41335	96244	29002	46453
53766	52875	15987	46962	67342	77592	57651	95508	80033	69828
90585	58955	53122	16025	84299	53310	67380	84249	25348	04332
32001	96293	37203	64516	51530	37069	40261	61374	05815	06714
62606	64324	46354	72157	67248	20135	49804	09226	64419	29457
10078	28073	85389	50324	14500	15562	64165	06125	71353	77669
91561	46145	24177	15294	10061	98124	75732	00815	83452	97355
13091	98112	53959	79607	52244	63303	10413	63839	74762	50289

Examples

1. Fortune publishes data on sales, profits, assets, stockholders' equity, market value, and earnings per share for the 500 largest U.S. industrial corporations (Fortune 500, 2003). Assume that you want to select a simple random sample of 10 corporations from the Fortune 500 list. Use the last three digits in column 9 of Table 1, beginning with 554. Read down the column and identify the numbers of the 10 corporations that would be selected.

2. The 10 most active securities on the New York (NYSE), Nasdaq, and American (AMEX) exchanges with market caps greater than $500 million are as follows (The Wall Street Journal, February 21, 2003):

Applied Materials	Nasdaq 100
Cisco Systems	Nextel
Intel	Oracle
Lucent Technologies	SPDR
Microsoft	Sun Microsystems

   a. Assume that a random sample of five securities will be selected for an in-depth study of trading behavior. Beginning with the first random digit in Table 7.1 and reading down the column, use the single-digit random numbers to select a simple random sample of five securities to be used in this study.
   b. According to the Notes and Comments information, how many different simple random samples of size 5 can be selected from the list of 10 securities?

3. Indicate whether the following populations should be considered finite or infinite.
   a. All registered voters in the state of California.
   b. All television sets that could be produced by the Allentown, Pennsylvania, plant of the TV-M Company.
   c. All orders that could be processed by a mail-order firm.
   d. All emergency telephone calls that could come into a local police station.
   e. All components that Fibercon, Inc., produced on the second shift on May 17.

Point Estimation

The sample mean $\bar{x}$ → the point estimator of the population mean μ

$$\bar{x} = \frac{\sum x}{n}$$

The sample standard deviation s → the point estimator of the population standard deviation σ

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

The sample proportion $\bar{p}$ → the point estimator of the population proportion p

$$\bar{p} = \frac{x}{n}$$

Examples

1. The following data are from a simple random sample: 5 8 10 7 10 14
   a. What is the point estimate of the population mean?
   b. What is the point estimate of the population standard deviation?

2. A sample of 50 fortune 500 companies (Fortune, April 14, 2003) showed 5 were based in New York, 6 in California, 2 in Minnesota, and 1 in Wisconsin.
    a. Develop an estimate of the proportion of Fortune 500 companies based in New York.
    b. Develop an estimate of the number of Fortune 500 companies based in Minnesota.
    c. Develop an estimate of the proportion of Fortune 500 companies that are not based in these 4 states.

Introduction to Sampling Distributions

If we consider the process of selecting a simple random sample as an experiment, the sample mean $\bar{x}$ is a numerical description of the outcome of the experiment. Thus, the sample mean $\bar{x}$ is a random variable. As a result, just like other random variables, $\bar{x}$ has a mean or expected value, a standard deviation, and a probability distribution. The probability distribution of $\bar{x}$ is called the sampling distribution of $\bar{x}$.