Chapter Thirteen

Basic Sampling Issues

LEARNING OBJECTIVES

1. Understand the concept of sampling.
2. Learn the steps in developing a sampling plan.
3. Understand the concepts of sampling error and nonsampling error.
4. Understand the differences between probability samples and nonprobability samples.
5. Understand sampling implications of surveying over the Internet.

The Concept of Sampling

Population

The entire group of people about whom information is needed; also called the *universe* or *population of interest*.

Sampling

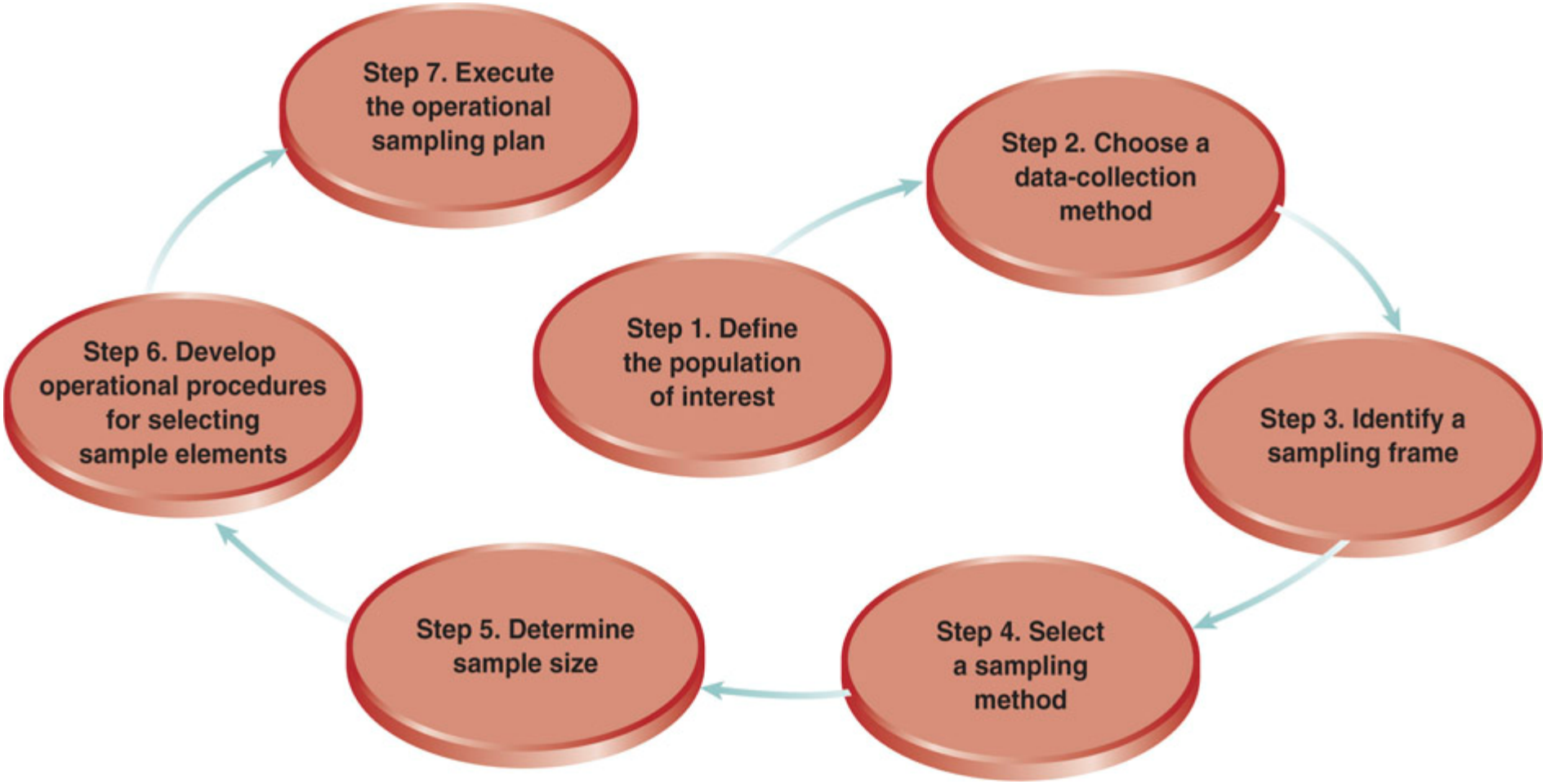
The process of obtaining information from a subset of a larger group.

Census vs. Sample

Census: Data collection from or about every member of the population of interest. Also called *canvassing* the population by asking *everyone* a set of questions.

Sample: A subset of all the members of a population of interest. Hopefully that subset is statistically representative of the general population.

Developing a Sampling Plan



Key Terms & Definitions

Developing a Sampling Plan

1. Define the Target Population

Determine the characteristics of those you are interested in studying. Determine which group of people or entities about which you want to learn more.

2. Choose the Data Collection Method

Determine how you collect the sample - such as mail, Internet, telephone, mall intercept, etc.

3. Select the Sample Frame

A list of population elements from which units to be sampled can be selected.

Developing a Sampling Plan

4. Select the Sampling Method

Determine how you will get the sample list through probability or non-probability methods.

5. Determine the Sample Size

What is the level of accuracy you want to achieve; the time and money you have to do the survey, and the data collection method?

6. Determine Operational Procedures

This is the plan of how to go about actually choosing and interviewing the respondents.

7. Execute the Sampling Plan

Field workers must be trained to execute the sampling plan properly.

Example

Develop a sampling plan for examining undergraduate business students' attitudes toward Internet advertising.

1.Determine the population: All undergraduate students that attend the university.

2.Determine data collection method: Email or phone Survey

3.Determine the sampling frame: A list of all undergraduate students divided by rank or class level obtained from the registrar's office

4.Determine sampling method: A quota sample based on gender and class standing

Example (cont'd)

5. Determine sample size: Depending upon statistical (level of error) requirements, a number will be chosen that is representative of the population of interest.
6. Plan for Selecting Sample Elements Since a quota sample has been chosen, a proportional representation of the population of male and females, and freshmen, sophomores, juniors, seniors, and graduate students will be determined. From each subset, an adequate number of students will be chosen.
7. Execution of Sampling Plan Telephone the chosen students and collect data using a questionnaire.

1. Defining the Population of Interest

EXHIBIT 13.2 Some Bases for Defining the Population of Interest

Geographic Area	What geographic area is to be sampled? This is usually a question of the client's scope of operation. The area could be a city, a county, a metropolitan area, a state, a group of states, the entire United States, or a number of countries.
Demographics	Given the objectives of the research and the target market for the product, whose opinions, reactions, and so on are relevant? For example, does the sampling plan require information from women over 18, women 18–34, or women 18–34 with household incomes over $35,000 per year who work and who have preschool children?
Usage	In addition to geographic area and/or demographics, the population of interest frequently is defined in terms of some product or service use requirement. This is usually stated in terms of use versus nonuse or use of some quantity of the product or service over a specified period of time. The following examples of use screening questions illustrate the point: ■ Do you drink five or more cans, bottles, or glasses of diet soft drinks in a typical week? ■ Have you traveled to Europe for vacation or business purposes in the past two years? ■ Have you or has anyone in your immediate family been in a hospital for an overnight or extended stay in the past two years?
Awareness	The researcher may be interested in surveying those individuals who are aware of the company's advertising, to explore what the advertising communicated about the characteristics of the product or service.

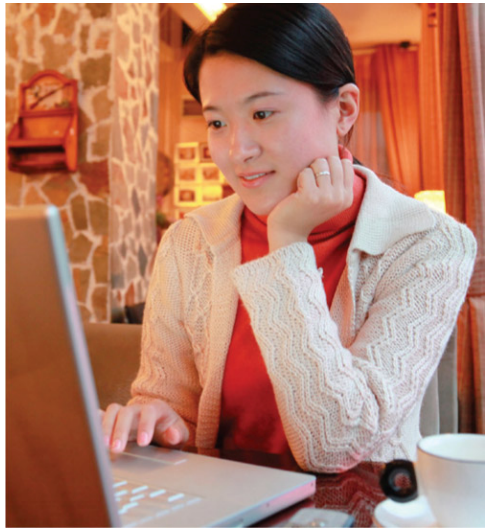
Key Terms & Definitions

2. Choose a Data Collection Method

This was discussed in the chapter on questionnaire design, also.

The sample plan will depend on the data collection method, too.

The main methods are:



- Personal, such as mall intercepts, etc.
- Telephone
- Mail or other self-administered
- Internet, blended samples

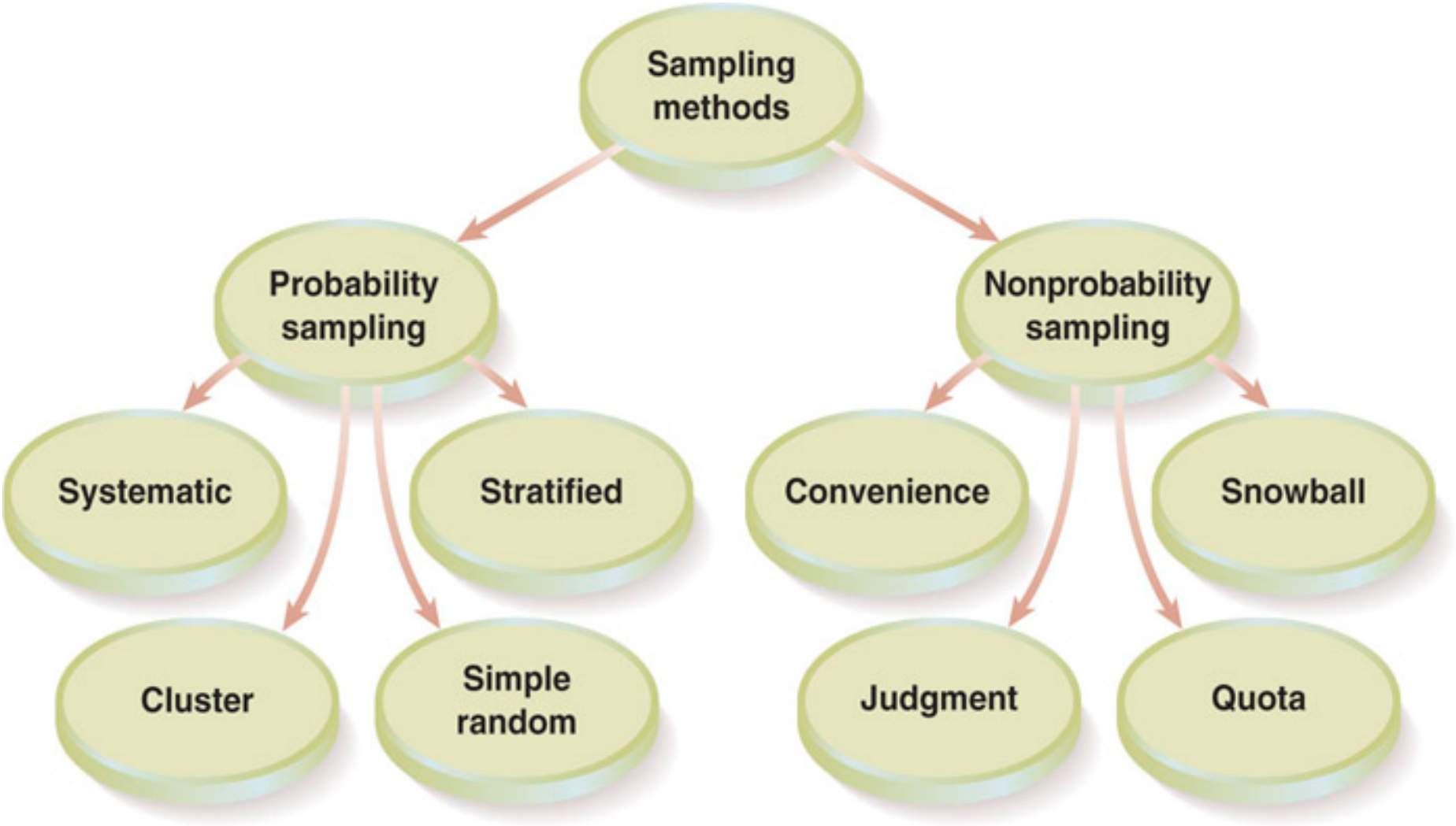
3. Identify the Sampling Frame

This is **a list of the members or elements of the population** from which units to be sampled are to be selected.

There is seldom a perfect correspondence between the sample frame and the population of interest.

Explain how a city's telephone book could be used as a sample frame of people in the city with telephones. What are the advantages and disadvantages of using that frame?

4. Sampling Methods



Probability Sampling

Simple Random

A sample selected by assigning a number to every element of the population and then using some method for randomly selecting elements to be in the sample such as random digit dialing.

$$\text{Probability of selection} = \frac{\text{Sample size}}{\text{Population size}}$$

Systematic

A sample in which the entire population is numbered and elements are selected using a skip interval – every "*nth*" name is selected.

$$\text{Skip interval} = \frac{\text{Population size}}{\text{Sample size}}$$

Simple Random

- Put 100 numbered bingo balls into a bowl (this is the population N).
- Select 10 balls from the bowl without looking(this is your sample n).

Simple random sample

The names of 200 patrons of a casino are drawn from a list of visitors for the last month and a questionnaire is administered to them.

Systematic

Example

- **American National Bank has 1,000 customers. The manager wishes to draw a sample of 100 customers.**
- **How would this be done using systematic sampling?**

Example

- Using systematic sampling, a skip interval of 10 would be determined (1000/10). Every tenth person would be chosen from the list to be included in the sample.

- Another technique that would yield a more randomly selected sample would be to select every customer whose Social Security number ended in 7, or 4, or whatever. That would be a 10% sample. The last digit of a Social Security number is a random number.

Probability Sampling

Stratified

A sample that is forced to be more representative through simple random sampling of mutually exclusive and exhaustive subsets either proportionally or disproportionally. Good for data that are not normally distributed.

Stratified Sampling Allocations

Proportional Allocation:

Sampling in which the number of elements selected from a stratum is directly proportional to the size of the stratum relative to the size of the population.

Disproportional, or Optimal, Allocation:

Sampling in which the number of elements taken from a given stratum is proportional to the relative size of the stratum and the standard deviation of the characteristic under consideration.

stratified sample

- Purpose: determine the GPA of college students across the U.S.
- Population: 21 million college students.
- Sample: random sampling 4,000 students.
  - English major: 560
  - Science major: 1135
  - Computer science: 800
  - Engineering: 1090
  - Math: 415

Stratified sample

- Sample: random sampling 4,000 students.
  - English major: 560 (14%)
  - Science major: 1135 (28.4%)
  - Computer science: 800 (20%)
  - Engineering: 1090 (27.3%)
  - Math: 415 (10.4%)
- Population
  - English major: 12%
  - Science major: 28%
  - Computer science: 24%
  - Engineering: 21%
  - Math: 15%

Stratified Sample

- Population
  - English major: 12%
  - Science major: 28%
  - Computer science: 24%
  - Engineering: 21%
  - Math: 15%
- Resample
  - English major: 12% (480)
  - Science major: 28% (1120)
  - Computer science: 24% (960)
  - Engineering: 21% (840)
  - Math: 15% (600)

Stratified sample

- A researcher invokes a pilot sample and finds that respondents from households with less than $50,000 annual income respond very differently than respondents from households with greater than $50,000 annual income, with regard to the key survey questions.
- In addition, the researcher's preliminary results show greater variance among respondents in the over $50,000 households.
- Given the preceding, which probability sampling method should the researcher invoke and why?

Stratified sample

- When a key demographic factor is related directly to the key survey questions and objectives, and there is great variance among respondents across the demographic factors, a stratified random sample should be used.
  - Two strata: <=$50000, and >$50,000

- Can use disproportionate stratified sampling:
  - More samples from household income>$50,000

Probability Sampling

Stratified

A sample that is forced to be more representative through simple random sampling of mutually exclusive and exhaustive subsets either proportionally or disproportionally. Good for data that are not normally distributed.

Cluster

A sample in which the sampling units are selected from a number of small geographic areas to reduce data collection costs.

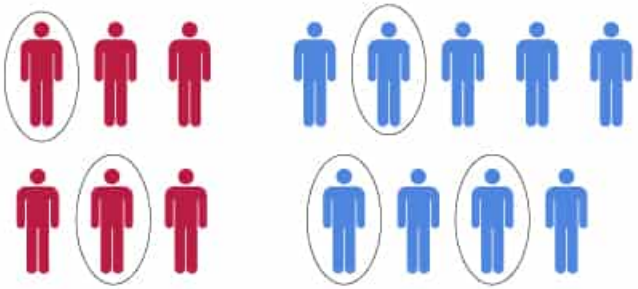
Stratified versus Cluster

- For a stratified random sample, a population is divided into stratum, or sub-populations, before sampling.

- At first glance, the two techniques seem very similar. However, in cluster sampling the actual **cluster is the sampling unit.**

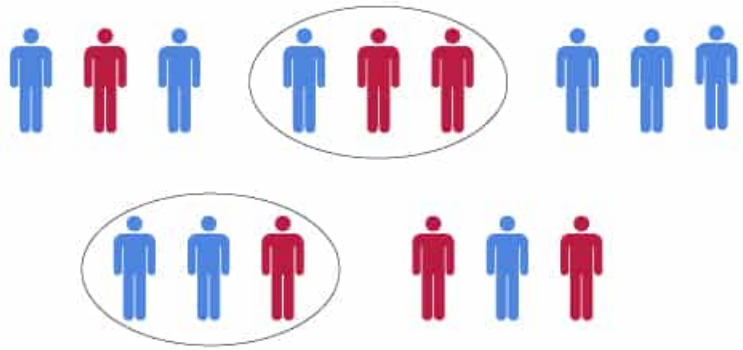
- in stratified sampling, analysis is done on **elements within each strata**.
- In cluster sampling, a researcher will only study selected clusters; with stratified sampling, a random sample is drawn from each strata.

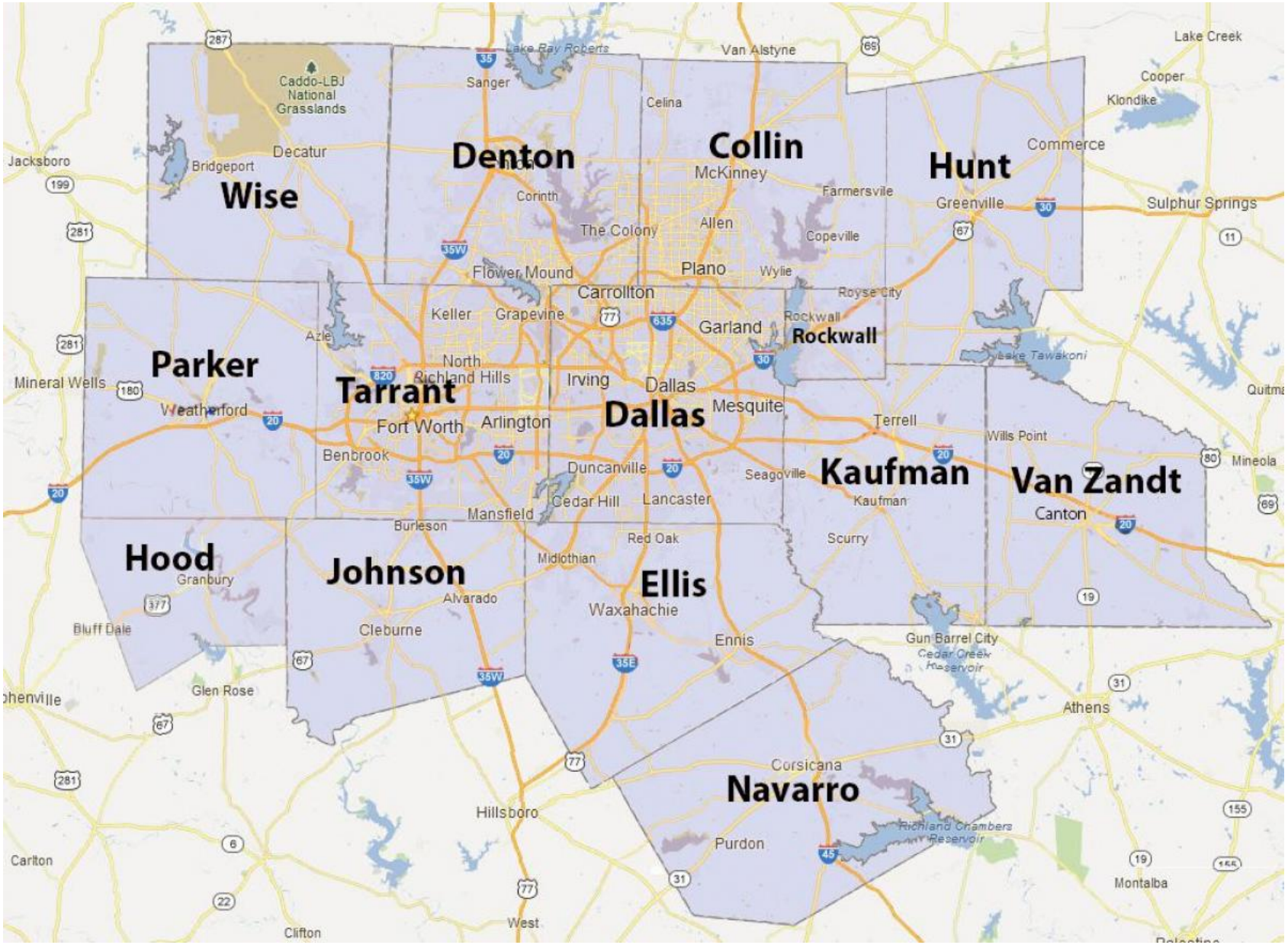
Stratified random sampling

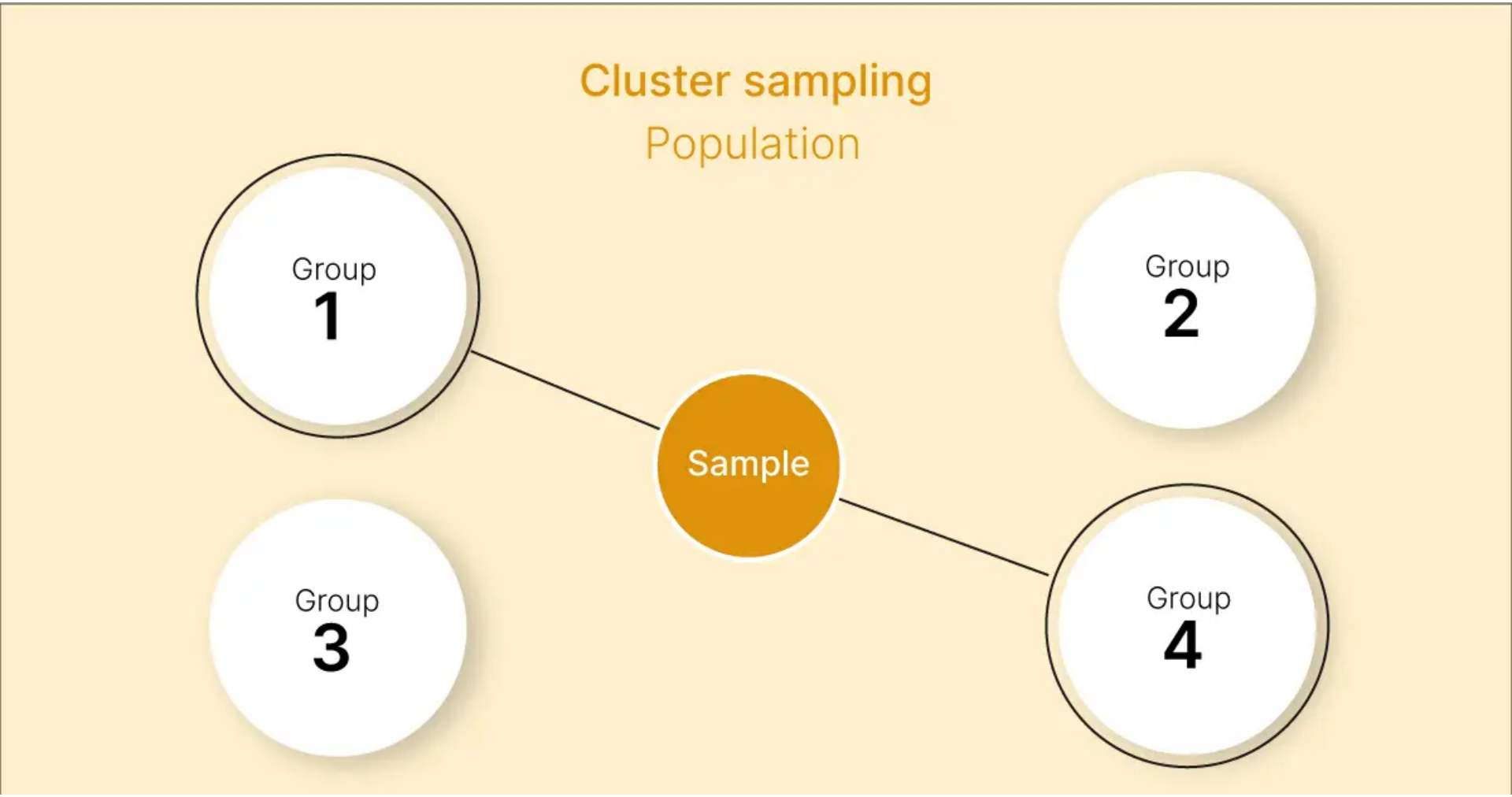


VS

Cluster sampling




Wise
Denton
Collin
Hunt
Parker
Tarrant
Dallas
Rockwall
Kaufman
Van Zandt
Hood
Johnson
Ellis
Navarro
Caddo-LBJ National Grasslands
Lake Ray Roberts
Lake Tawakoni
Richland Chambers Reservoir
Cedar Creek Reservoir
Jacksboro
Bridgeport
Decatur
Sanger
Van Alstyne
Celina
McKinney
Corinth
The Colony
Allen
Farmersville
Copeville
Commerce
Greenville
Lake Creek
Cooper
Klondike
Sulphur Springs
Flower Mound
Plano
Wylie
Carrollton
Royse City
Keller
Grapevine
Garland
Rockwall
Azle
North Richland Hills
Irving
Dallas
Mesquite
Terrell
Wills Point
Mineral Wells
Weatherford
Fort Worth
Arlington
Benbrook
Duncanville
Seagoville
Kaufman
Canton
Mineola
Cedar Hill
Lancaster
Mansfield
Burleson
Red Oak
Scurry
Midlothian
Granbury
Alvarado
Waxahachie
Ennis
Cleburne
Bluff Dale
Glen Rose
Gun Barrel City
Athens
Corsicana
Hillsboro
Purdon
Montalba
Carlton
Clifton
West


Cluster sampling
Population
Group 1
Group 2
Sample
Group 3
Group 4

Key Point Difference between Stratified Sampling and Cluster Sampling

Stratified Sampling

1. Elements within each stratum are sampled.
2. From each stratum, a random sample is selected.
3. Aim: to increase precision to reduce error.

Cluster Sampling

1. Only selected clusters are sampled.
2. Cluster/group is considered a sampling unit.
3. Aim: reduce cost and increase the efficiency of sampling.

Sampling and Nonsampling Error

Sampling Error

Error that occurs because the sample selected is not perfectly representative of the population.

Nonsampling Error

All error other than sampling error - also called "measurement error."

Nonprobability Sampling

Convenience

A sample based on using people who are easily accessible – such as mall intercepts or other high traffic locations.

Judgment

A sample in which the selection criteria are based on the researcher's personal judgment about the representativeness of the population under study. The researcher selects who should be in the study.

Nonprobability Sampling

Quota

A sample in which quotas, based on demographic or classification factors selected by the researcher, are established for population subgroups.

Snowball

A sample in which additional respondents are selected based on referrals from initial respondents.

Convenience Example

A radio talk show host invites listeners to call in and vote yes or no on whether handguns should be banned.

Judgement Sampling Example

- What it takes for American youths to graduate from high school by age 14, instead of the typical graduation age of 18 years old.

- A random sample? No
  - That includes a significant amount of youths that are on track to graduate at the traditional age of 18 years old.

- Focus only on the members of the population that fit the criteria.

- Judgment sampling is the only viable option.

Quota Example

A dog food manufacturer wants to test a new dog food.

It decides to select:

1. **100 dog owners who feed their dog's canned food,**
2. **100 who feed their dog's dry food,**
3. **and 100 who feed their dog's semi moist food.**

Snowball Sampling

- Looking for individuals who have been a victim of a particular crime
- A victim support network in the area.
- There may be no other way to obtain the respondent's names.
- The danger:
  - The sample may not be a good cross section
  - Respondents may be reluctant to give referrals.

Snowball Sampling

- Doing a survey of snow skiers residing in South Louisiana
- Purpose: estimating the feasibility of offering discounted snow ski rental equipment.
- The idea: gauge the need, interest and price expectations of these snow skiers as they head to Colorado or some other snow skiing destination.
- What type of sample would be needed?

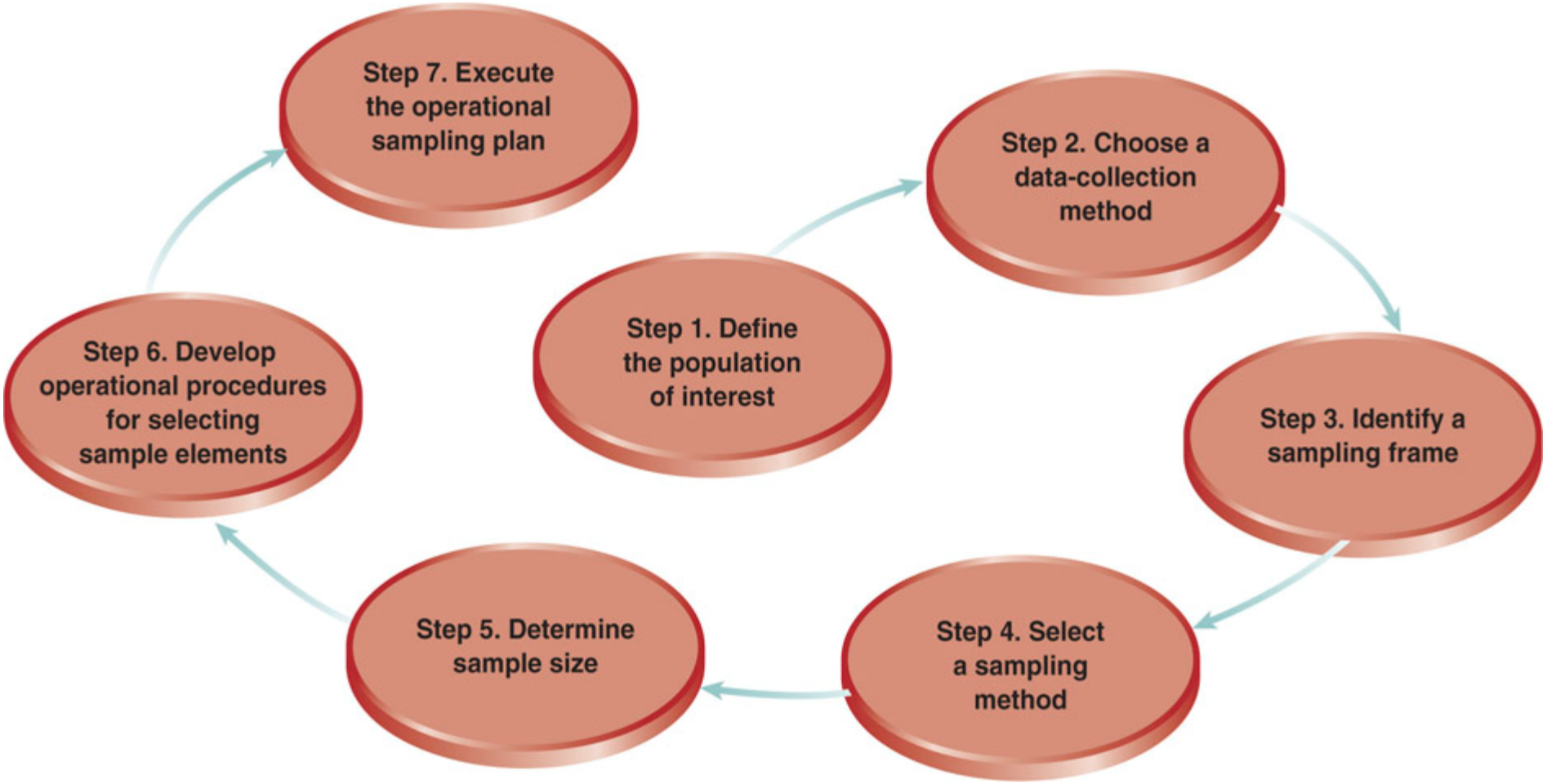
Snowball Sampling

- It is plausible to assume that there will a low incidence of snow skiers in South Louisiana. Instead of surveying at random, the researcher should resort to a referral or snowball sampling technique.

Systematic sample, cluster sample, and stratified sample

- A systematic sampling procedure uses **a skip interval** which is calculated by dividing the total population by the sample size.
- A cluster sampling method considers groups of people in a particular geographical region.
  - For example, a cluster sample could be used in retail trade analysis where a geographic area is clustered on a specific criterion (street boundaries). A specific number of clusters are randomly chosen. Then respondents are chosen randomly within each cluster either by a simple random scheme or a systematic sampling scheme.
- A stratified sampling procedure divides a population by specific strata (some demographic characteristic pertinent to the population of interest) then people are chosen randomly within each stratum, usually proportionate to the total number of people in each stratum.

Developing a Sampling Plan



Key Terms & Definitions

5. Determine the Sample Size

Determine the Sample Size

Determining the sample size will be based on factors such as the level of accuracy you want to achieve, the time and money you have to do the survey, and on the sampling collection method.

The actual size could be determined by a formula or by some rule of thumb based on experience.

6. Develop Operational Procedures

Select Sample Units

Your operational plan to conduct the probability or non- probability sampling. Determine the phases of the sample selection process. Multi-stage sampling involves combining sampling methods.

For probability sampling, these instructions need to be detailed, clear, and unambiguous and should eliminate any interviewer discretion regarding the selection of specific sample elements.

7. Execute the Plan

Conduct Fieldwork

The execution phase of the research. Administering the questionnaire - sending the mailers, making the phone calls, conducting the mall intercepts, etc.

As mentioned in the last slide, if a probability sample is used, the instructions need to be followed to the letter. Good training and supervision of fieldwork personnel (if they are used) is of utmost importance.

Internet Sampling

Pros

- Target respondents can complete the survey at their convenience.
- Data collection is inexpensive.
- Survey software can facilitate the data collection process.
- The survey can be completed quickly.

Cons

- Sample might not be representative of the population.
- You cannot always be sure who is completing the survey.
- Maintaining respondent confidentially can be a challenge.
- Data security issues can be difficult to manage.

Chapter Fourteen

Sample Size Determination

LEARNING OBJECTIVES

1. Discover methods for determining sample size.
2. Gain an appreciation of a normal distribution.
3. Understand population, sample, and sampling distributions.
4. Understand how to compute the sampling distribution of the mean.
5. Learn how to determine sample size.
6. Understand how to determine statistical power.

Sample Size for Probability Sampling

Census:

- Population canvas - not really a “sample”
- Asking the entire population

Budget Available:

- how much can we afford?
- budget=$20, 000, each survey=$20

Rule of Thumb:

- Is there some convention we can apply?
- Experience based.

Sample Size for Probability Sampling

Number of Subgroups Analyzed:

- 50% of male and 50% of female
- Do we get enough size for each subgroup?

Traditional Statistical Methods:

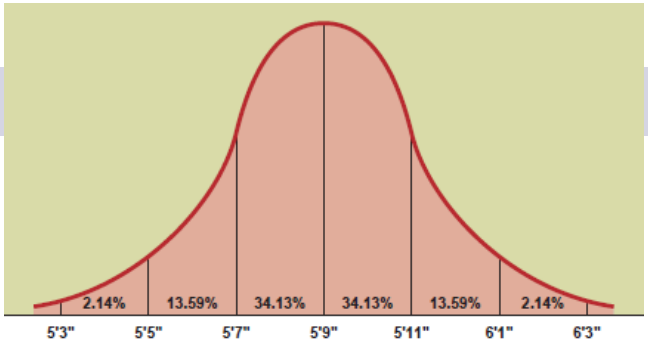
Variance, standard deviation, and confidence interval play a key role.

The Normal Distribution

Central Limit Theorem:

The idea that a distribution of a large number of sample means or sample proportions will approximate a normal distribution - regardless of the distribution of the population from which they were drawn.

Normal Distribution:

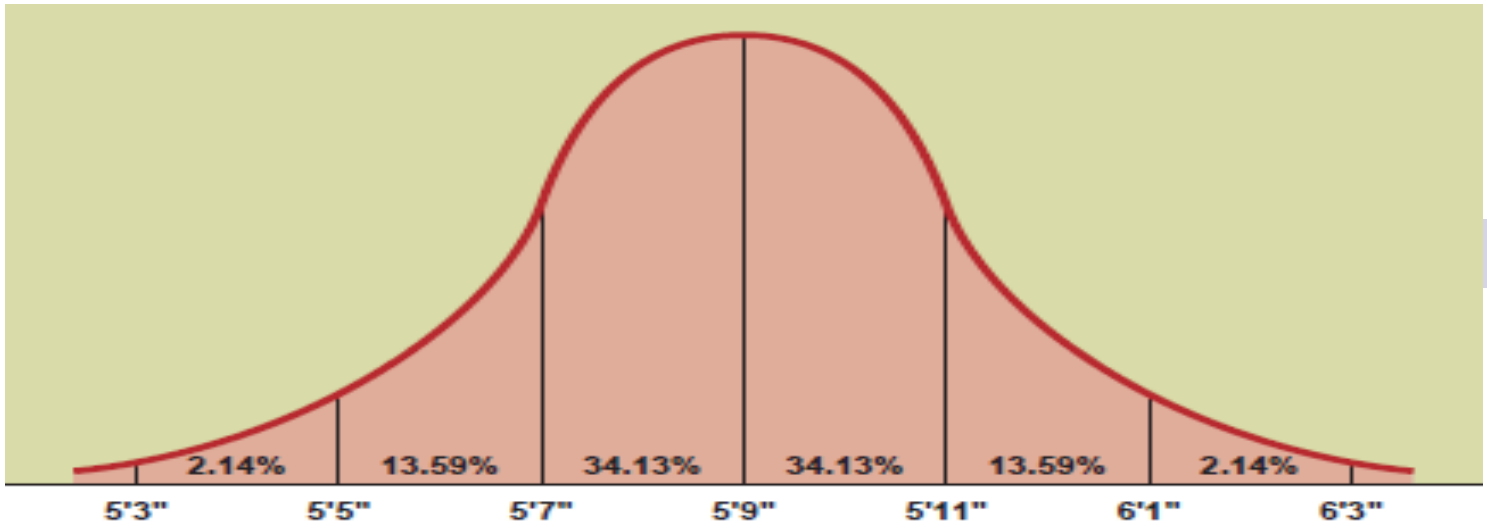


- Bell shaped
- Symmetrical about the mean.
- The mean, median, and mode are equal.
- Finally, about 68% of the observations are within one standard deviation plus/minus of the mean, 96% are within two standard deviations, and 99+% are within three standard deviations of the mean respectively.

The Normal Distribution

Proportionate Properties of the Normal Distribution:

A feature that the number of observations falling between the mean and a given number of standard deviations from the mean is the same for all normal distributions.



Question

- If we were making inferences from a single sample, we would expect that there would be a ____ percent probability that the sample mean or proportion generated from our sample results would be within two standard errors of the true population mean.

A. 68.26%
B. 99.74%
C. 97.5%
D. 95.44%
E. 98.00%

The Normal Distribution

Standard Deviation:

The measure of dispersion calculated by subtracting the mean of the series from each value in a series, squaring each result, summing the results, dividing the sum by the number of observations minus 1 and finally taking the square root of this value.

$$\text{Standard Deviation} = \sqrt{\frac{\sum (X_1 - \overline{X})^2}{(N-1)}}$$

Population and Sampling Distribution

Population Distribution:

The frequency distribution of all the elements of a population.

μ σ

Sampling Distribution:

The frequency distribution of all the elements of an individual sample.

$\bar{X}$ S

Sampling Distribution of the Mean

Sampling Distribution of the Mean:

The theoretical frequency distribution of the means of all possible samples of a given size drawn from a particular population; it is normally distributed.

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Standard Error of the Mean

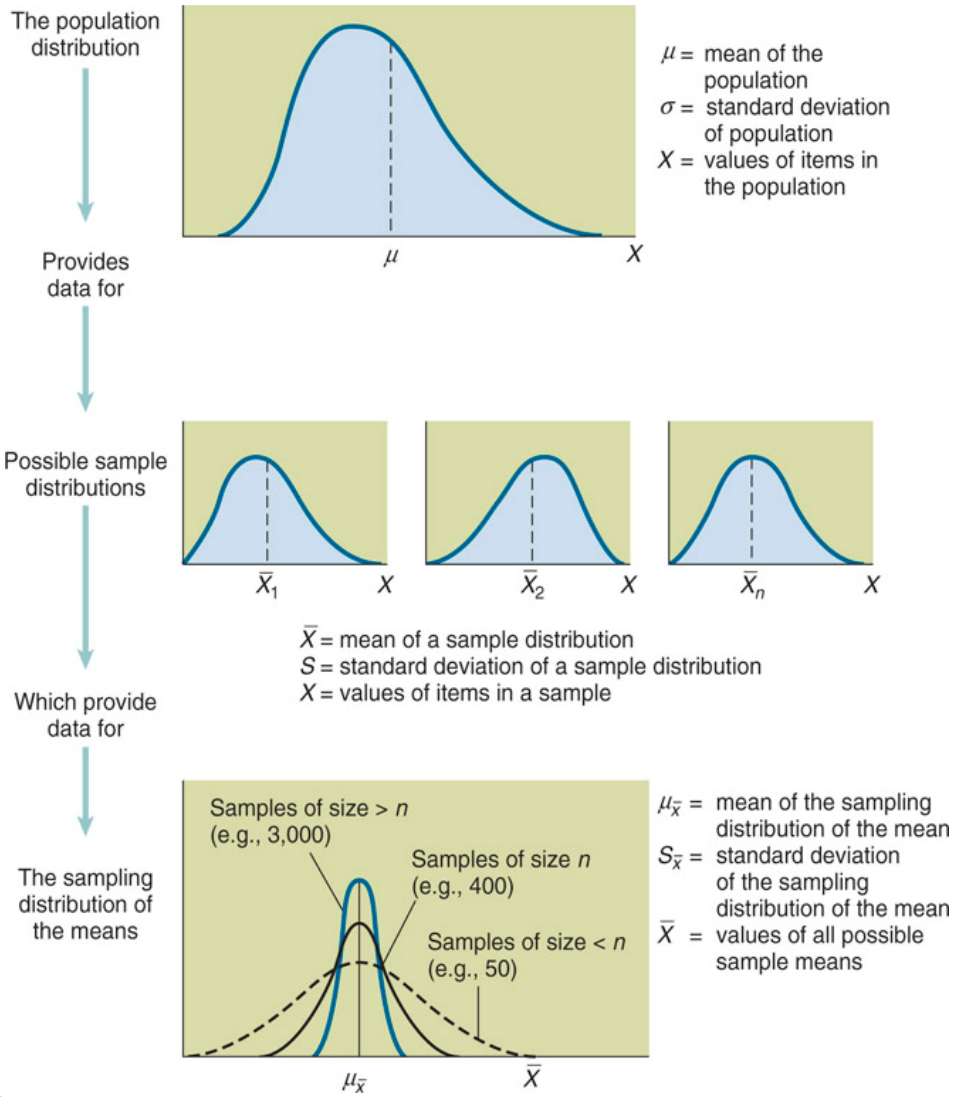
Standard Error of the Mean:

Standard deviation of a distribution of sample means.

EXHIBIT 14.4 Notation for Means and Standard Deviations of Various Distributions

Distribution	Mean	Standard Deviation
Population	μ	σ
Sample	$\overline{X}$	S
Sampling	$\mu_{\bar{x}} = \mu$	$S_{\bar{x}}$

Relationships of the Three Basic Types of Distribution



Standard Error of the Mean

- 1000 consumers:
  - number of time eating in a restaurant in last month
- Draw a size of 200 from the 1000 consumers. Draw 100 times.
- Each time, we get a mean number of a 200 customers, the mean is : $\bar{x}_1, \bar{x}_2, \bar{x}_3, ..., \bar{x}_{100}$
-
- The standard deviation of the above sample means:
  - standard error of the mean
  - Following a normal distribution

Example

Standard error of the mean is:

a. a measure of dispersion of a large population.

b. a measure of dispersion of large sample.

c. the standard deviation of a distribution of sample means.

d. the error in the computation of population standard deviation.

e. none of these

Sampling Distribution of the Proportion

Sampling Distribution of the Proportion:

The relative frequency distribution of the sample proportions of many random samples of a given size drawn from a particular population; it is normally distributed.

$$S_p = \sqrt{\frac{P(1-P)}{n}}$$

where S_p = standard error of sampling distribution of proportion

P = estimate of population proportion

n = sample size

Sampling Distribution of the Proportion

•Estimating the percentage of all adults who have accessed Twitter in the past 90 days.

•Select 1,000 random samples of size 200 from the population of all adults.

•Compute the proportion of all adults who have accessed Twitter in the past 90 days for all 1,000 samples.

•The mean proportion for all possible samples is equal to the

$$S_p = \sqrt{\frac{P(1 - P)}{n}}$$

where S_p = standard error of sampling distribution of proportion

P = estimate of population proportion

n = sample size

Sampling Distribution of the Proportion

- Each sample of size 200 provides the proportion of all adults who have accessed Twitter in the past 90 days.

$$P_1, P_2, P_3, \ldots, P_{1000}$$

- The mean proportion for all possible samples is equal to the population proportion.
- $Mean(P_1, P_2, P_3, \ldots, P_{1000}) = P$

$$n = 200$$

$$S_p = \sqrt{\frac{P(1 - P)}{n}}$$

where S_p = standard error of sampling distribution of proportion

P = estimate of population proportion

n = sample size

Determining the Sampling Size

Problems Involving Mean:

$$n = \frac{Z^2\sigma^2}{E^2}$$

where Z = level of confidence expressed in standard errors

σ = population standard deviation

E = acceptable amount of sampling error

Three pieces of information are needed to compute the sample size required:

1. The acceptable or allowable level of sampling error E.
2. The acceptable level of confidence Z. In other words, how confident does the researcher want to be that the specified confidence interval includes the population mean?
3. An estimate of the population standard deviation σ.

Determining the Sampling Size

1. Allowable Sampling Error:

Amount of sampling error the researcher is willing to accept, E.

2. The Acceptable Level of Confidence:

How confident does the researcher want to be that an interval includes the population mean, Z.

3. Population Standard Deviation:

Standard deviation of a variable for the entire population, σ.

Example

- **You are in charge of planning a chili cook-off. You must make sure that there are plenty of samples for the patrons of the cook-off.**
- **The following standards have been set: a confidence level of 99 percent and an error of less than 2 ounces per cooking team. Last year's cook-off had a standard deviation in amount of chili cooked of 3 ounces. What is the necessary sample size?**

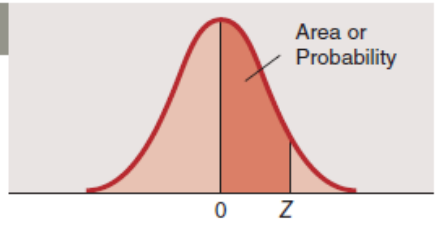
Answer

$$n = \frac{z^2\sigma^2}{E^2} = \frac{2.6^2 * 3^2}{2^2} = 15.21$$

- Z=2.6 for confidence level of 99 percent
  - i.e., 2.6 standard errors are required to take in 99% of the area under a normal curve. See appendix one in the end of text book-Exhibit 2.
- Thus, a sample of 16 is needed. A sample of 30 will allow more powerful analysis.
- .99/2=.495 ⟶ z=2.6
- Z=2.6, area=.4953*2=.9906
- Z=2.59, area=.4952*2=.9904
- Z=2.61, area=.4955*2=.9910

EXHIBIT 2 Standard Normal Distribution: *Z*-values

Entries in the table give the area under the curve between the mean and *Z* standard deviations above the mean. For example, for $Z = 1.25$, the area under the curve between the mean and *Z* is .3944.



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974

Determining the Sampling Size

Problems Involving Proportions:

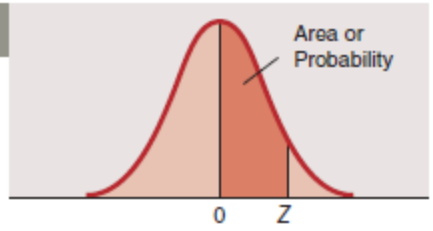
$$n = \frac{Z^2[p(1-p)]}{E^2}$$

Problems Involving Proportions

- A researcher wants to estimate the percentage of all adults that have used the Internet to seek pre-purchase information in the past 30 days, with a tolerable sampling error (E) of **0.03** and a confidence level of **97.5%**. If secondary data indicated that **25%** of all adults had used the Internet for such a purpose, what is the sample size?
- .975/2=.4875

EXHIBIT 2 Standard Normal Distribution: *Z*-values

Entries in the table give the area under the curve between the mean and *Z* standard deviations above the mean. For example, for $Z = 1.25$, the area under the curve between the mean and *Z* is .3944.



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936

Determining the Sampling Size

$$z = 2.24,$$

$$0.4875 * 2 = .975$$

Determining the Sampling Size

$$n = \frac{Z^2[p(1-p)]}{E^2}$$

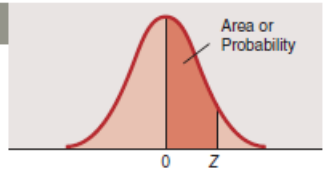
$$= \frac{2.24^2[.25*(1-.25)]}{0.03^2}$$

$$= 1046$$

Example

- **Assume that previous fast-food research has shown that 80 percent of the consumers like curly French fries. The researcher wishes to have an error of 6 percent or less and be 95 percent confident of an estimate to be made about curly French fry consumption from a survey. What sample size should be used for a simple random sample?**
- .95/2=.475

EXHIBIT 2 **Standard Normal Distribution: Z-values**



Entries in the table give the area under the curve between the mean and Z standard deviations above the mean. For example, for $Z = 1.25$, the area under the curve between the mean and Z is .3944.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4986	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Answer

$$n = \frac{z^2[p(1-p)]}{E^2} = \frac{1.96^2\left[.8(1-.8)\right]}{.06^2} = 170.7378$$

- Where z=1.96,p=.8, 1-p=1-.8, E=0.06
- Thus, a sample size of 171 subjects is needed.

How Many Contacts?

You need to contact more than n, since it's unlikely that everyone you contact will agree to answer the questionnaire. If you determine that you need 800 respondents and a typical response rate for your type of survey is 30%, then you would actually have to contact 2,667 people to get 800 completed surveys.

Number of Contacts = n/.30
Or
800/.3
Number of Contacts = 2,667

You must contact approximately 2,667 people (with 30% response rate), so you can expect 800 people to respond (the final sample size).

Example

- **A marketing researcher must determine how many telephone numbers she needs to order from a sample provider to complete a survey of ATM users.**
- **The goal is to complete 400 interviews with ATM users.**
- **From past experience, she estimates that 60 percent of the phone numbers provided will be working phone numbers. The estimated incidence rate (percentage of people contacted who are ATM users) is 43 percent. Finally, she estimates from previous surveys that 35 percent of the people contacted will agree to complete the survey. How many telephone numbers should she order?**

Answer

- 400 = (.6 * .43 * .35)*n. Thus, *n*= 4,430 telephone numbers.

Sample Questions

Which of the following statements about sample size is false?

a)There is no direct relationship between population size and the size of the sample required to estimate population parameters.

b)The larger the sample variance, the larger the sample size that is required.

c)The higher the level of confidence, the larger the sample size that is required.

d)The larger the population, the larger the sample size that is required.

e)All of these statements are false.