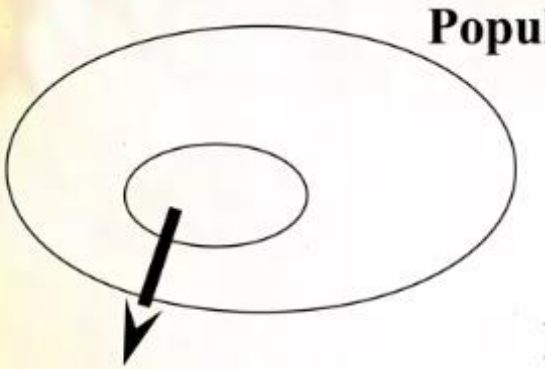
Sampling

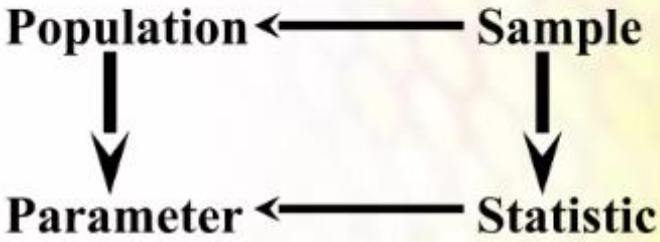
- The process of obtaining information from a subset (sample) of a larger group (population)
- The results for the sample are then used to make estimates of the larger group
- Faster and cheaper than asking the entire population
- Two keys
  1. Selecting the right people
     - Have to be selected scientifically so that they are representative of the population
  2. Selecting the right number of the right people
     - To minimize sampling errors I.e. choosing the wrong people by chance

SAMPLING

- Sample -- contacting a portion of the population (e.g., 10% or 25%)
  - best with a very large population (n)
  - easiest with a homogeneous population
- Census -- the entire population
  - most useful is the population ("n") is small
  - or the cost of making an error is high

Population Vs. Sample

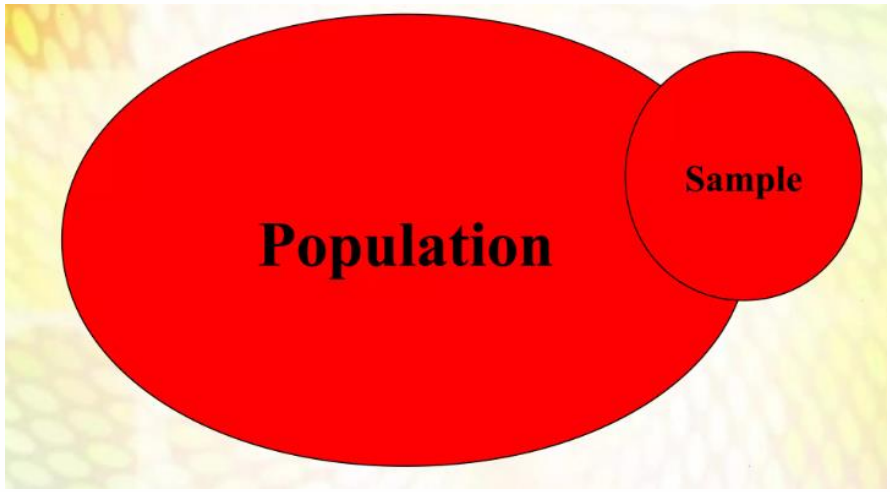


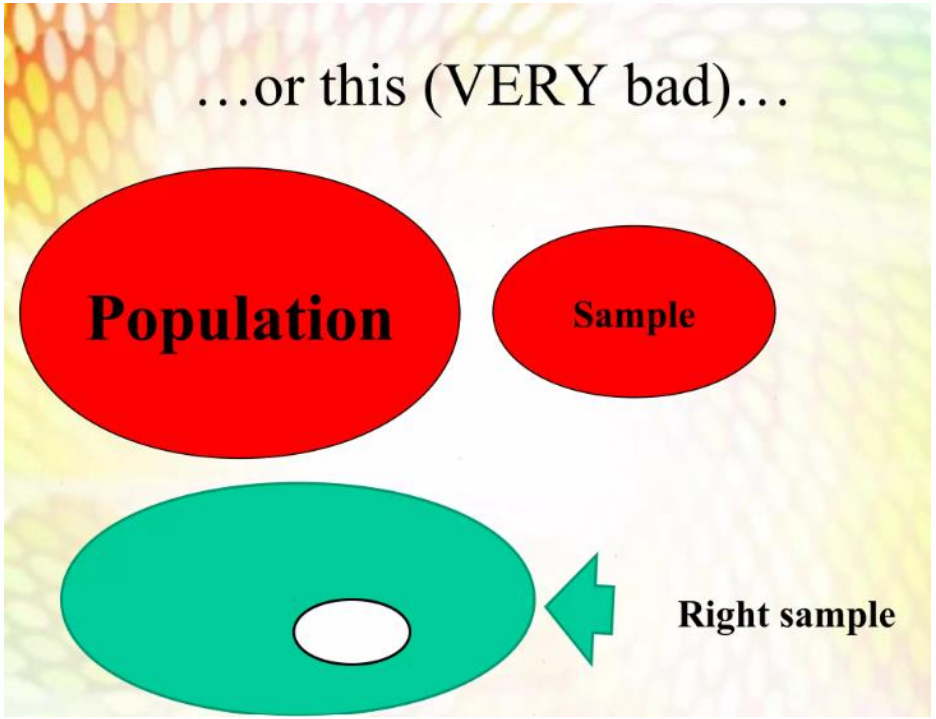


We measure the sample using statistics in order to draw inferences about the population and its parameters.

Characteristics of Good Samples

- Representative
- Accessible
- Low cost





Terminology

Population

- The entire group of people of interest from whom the researcher needs to obtain information.

Element (sampling unit)

- one unit from a population

Sampling

- The selection of a subset of the population

Sampling Frame

- Listing of population from which a sample is chosen

Census

- A polling of the entire population

Survey

- A polling of the sample

Terminology

Parameter

- The variable of interest

Statistic

- The information obtained from the sample about the parameter

Goal

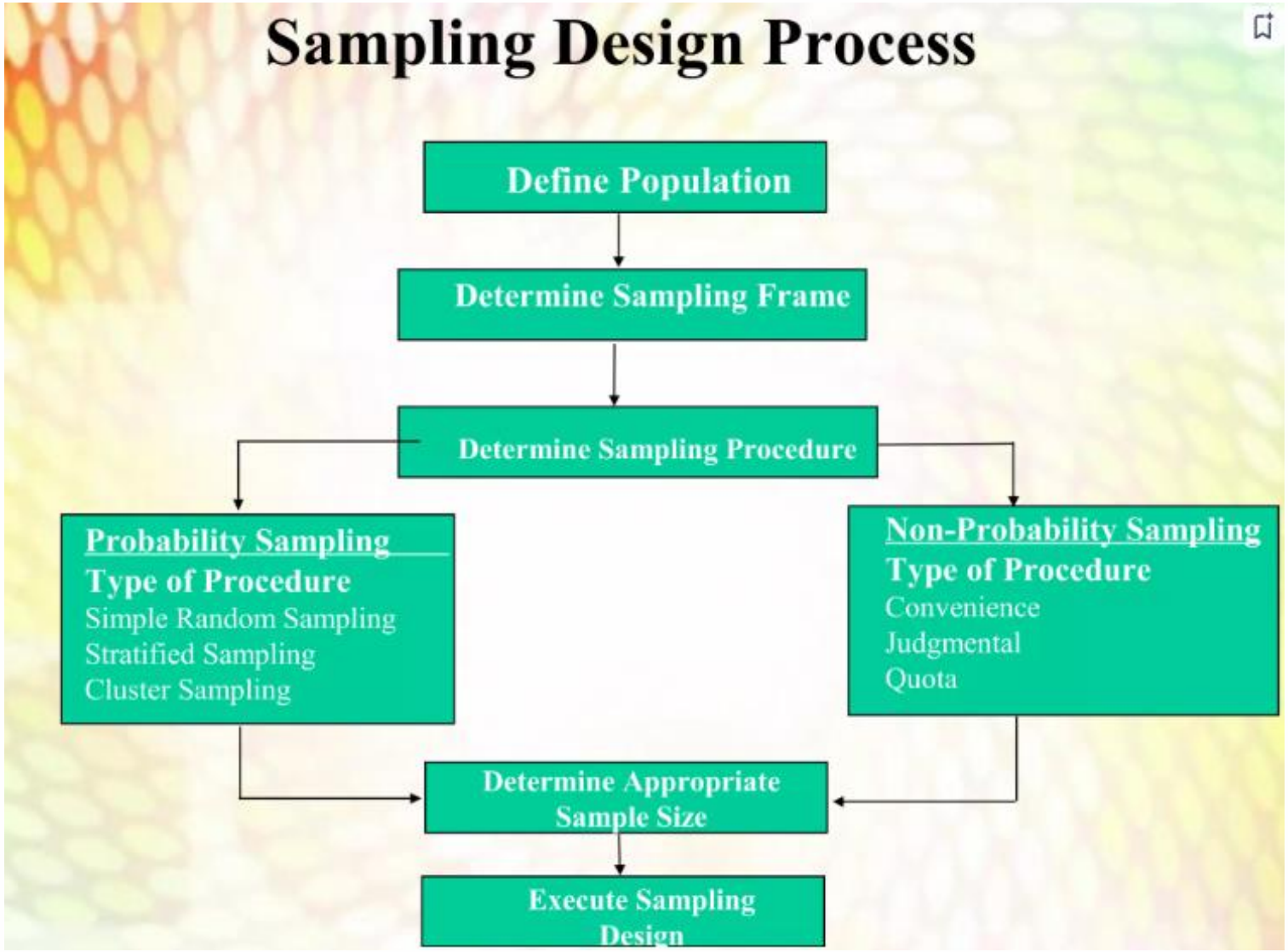
- To be able to make inferences about the population parameter from knowledge of the relevant statistic - to draw general conclusions about the entire body of units

Critical Assumption

- The sample chosen is representative of the population

Steps in Sampling Process

1. Define the population
2. Identify the sampling frame
3. Select a sampling design or procedure
4. Determine the sample size
5. Draw the sample


Sampling Design Process
Define Population
Determine Sampling Frame
Determine Sampling Procedure
Probability Sampling
Type of Procedure
Simple Random Sampling
Stratified Sampling
Cluster Sampling
Non-Probability Sampling
Type of Procedure
Convenience
Judgmental
Quota
Determine Appropriate Sample Size
Execute Sampling Design

Define the Target Population

- It addresses the question "Ideally, who do you want to survey?" i.e. those who have the information sought What are their characteristics. Who should be excluded?
  - age, gender, product use, those in industry
  - Geographic area

1. Define the Target Population

The Element	*Individuals* *families*
***S*ampling Unit….**	*individuals over 20* *families with 2 kids* *seminar groups at "new" university*
Extent	*individuals who have bought "one"* *families who eat fast food*
Timing	*bought over the last seven days*

The target population for a toy store can be defined as all households with children.

What's wrong with this definition?

2. Determine the Sampling Frame

- **Obtaining a "list" of population (how will you reach sample)**
  - Students who eat at McDonalds?
  - young people at random in the street?
  - phone book
  - students union listing
  - University mailing list
- **Procedures**
  - E.g. individuals who have spent two or more hours on the internet in the last week

2. Determine the Sampling Frame

Select "*sample units*"

- Individuals
- Household
- Streets
- Telephone numbers
- Companies

3. Selecting a Sampling Procedure

- **Probability sampling - equal chance of being included in the sample (random)**
  - simple random sampling
  - systematic sampling
  - stratified sampling
  - cluster sampling
- **Non-probability sampling - - unequal chance of being included in the sample (non-random)**
  - convenience sampling
  - judgement sampling
  - snowball sampling
  - quota sampling

3. Selecting a Sampling Procedure

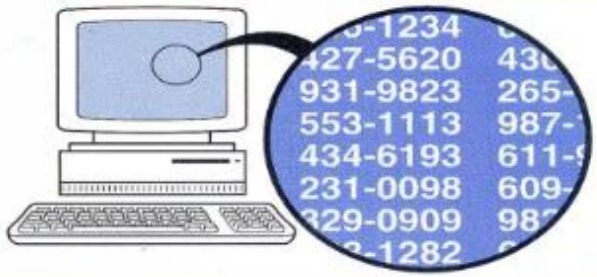
Probability Sampling

- An objective procedure in which the probability of selection is nonzero and is known in advance for each population unit.
- It is also called random sampling.
- Ensures information is obtained from a representative sample of the population
- Sampling error can be computed
- Survey results can be projected to the population
- More expensive than non-probability samples

3. Selecting a Sampling procedure

Simple Random Sampling (SRS)

- **Population members are selected directly from the sampling frame**
- **Equal probability of selection for every member (sample size/population size)**
- **400/10,000 = .04**
- **Use random number table or random number generator**



3. Selecting a Sampling procedure

Simple Random Sampling

- N = the number of cases in the sampling frame
- n = the number of cases in the sample
- ${}_{N}C_{n}$ = the number of combinations (subsets) of n from N
- f = n/N = the sampling fraction

3. Selecting a Sampling Design

Objective: To select *n* units out of *N* such that each ${}_NC_n$ has an equal chance of being selected

Procedure: Use a table of random numbers, a computer random number generator, or a mechanical device to select the sample

3. Selecting a Sampling Design

Systematic Sampling

- Order all units in the sampling frame based on some variable and number them from 1 to *N*
- Choose a random starting place from 1 to *N* and then sample every k units after that

systematic random sample

number the units in the population from 1 to N

N = 100

decide on the n (sample size) that you want or need

want n = 20

k = N/n = the interval size

N/n = 5

randomly select an integer between 1 to k

select a random number from 1-5: chose 4

then take every kth unit

start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

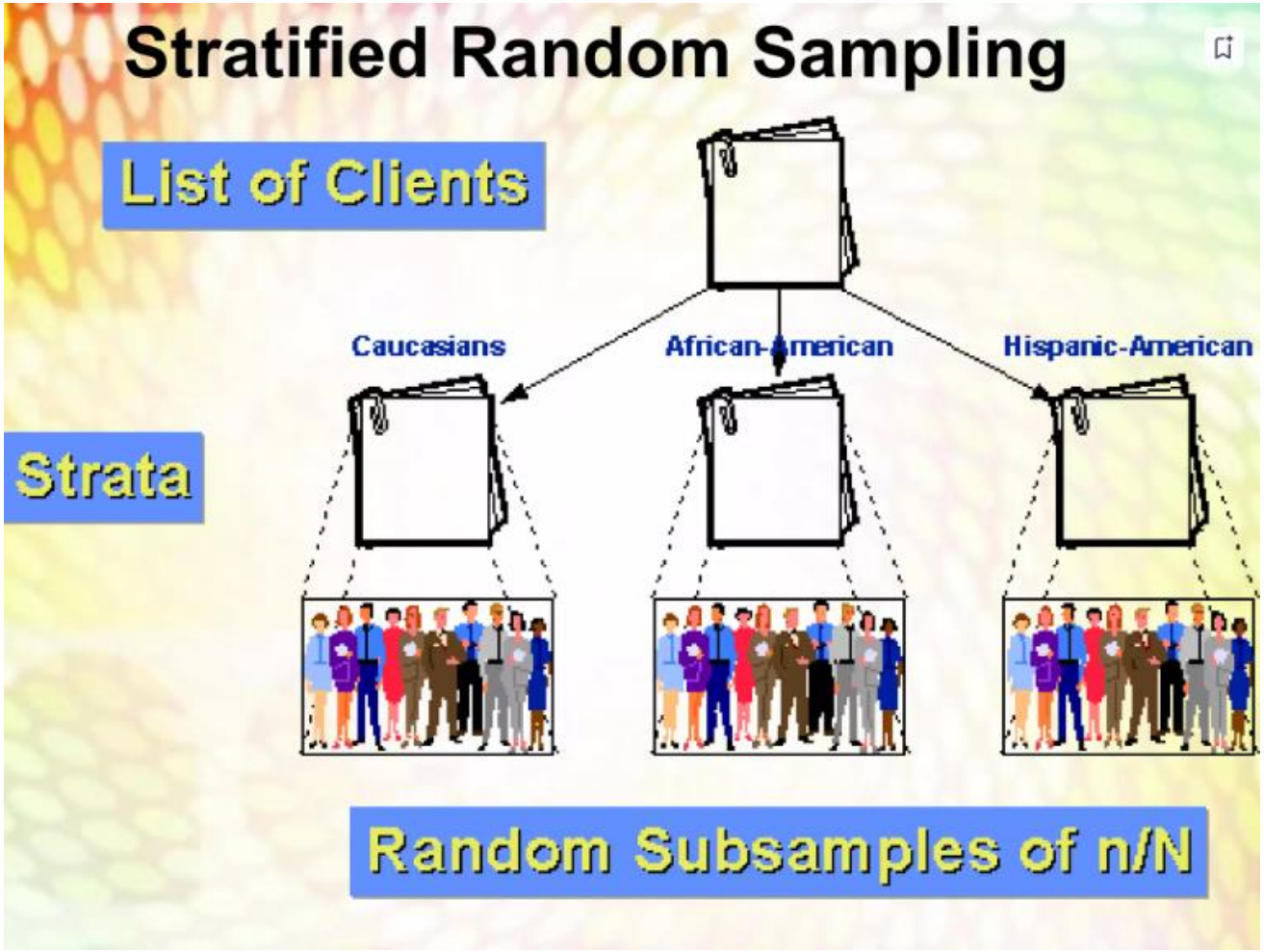
3. Selecting a Sampling Design

Stratified Sampling (I)

- The chosen sample is forced to contain units from each of the segments, or strata, of the population
  - equalizing "important" variables
    - year in school, geographic area, product use, etc.
- Steps:
  - Population is divided into mutually exclusive and exhaustive strata based on an appropriate population characteristic. (e.g. race, age, gender etc.)
  - Simple random samples are then drawn from each stratum.

Stratified Random Sampling

- Population is divided on the basis of characteristic of interest in the population e.g. male and female may have different consumption patterns
- Has a smaller sampling error than simple random sample since a source of variation is eliminated
- Ensures representativeness when proportional sampling used


Stratified Random Sampling
List of Clients
Caucasians
African-American
Hispanic-American
Strata
Random Subsamples of n/N

3. Selecting a Sampling Design
Stratified Sampling (II)

- **Direct Proportional Stratified Sampling**
  - **The sample size in each stratum is proportional to the stratum size in the population**
- **Disproportional Stratified Sampling**
  - **The sample size in each stratum is NOT proportional to the stratum size in the population**
  - **Used if**

1) some strata are too small

2) some strata are more important than others

3) some strata are more diversified than others

3. Selecting a Sampling Design

Cluster Sampling

- Clusters of population units are selected at random and then all or some randomly chosen units in the selected clusters are studied.

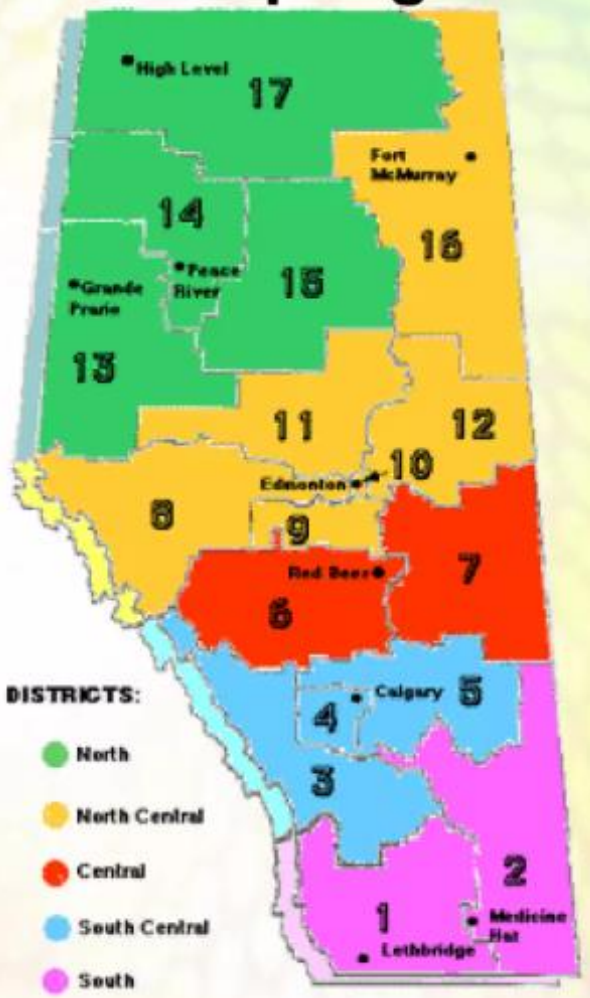
- Steps:
  - **Population is divided into mutually exclusive and exhaustive subgroups, or clusters. Ideally, each cluster adequately represents the population.**
  - **A simple random sample of a few clusters is selected.**
  - **All or some randomly chosen units in the selected clusters are studied.**

cluster or area random sampling

divide population into clusters (usually along geographic boundaries)

randomly sample clusters

measure units within sampled clusters



3. Selecting a Sampling Design

When to use stratified sampling

- **If primary research objective is to compare groups**
- **Using stratified sampling may reduce sampling errors**

When to use cluster sampling

- **If there are substantial fixed costs associated with each data collection location**
- **When there is a list of clusters but not of individual population members**

3. Selecting a Sampling Design

Non-Probability Sampling

- **Subjective procedure in which the probability of selection for some population units are zero or unknown before drawing the sample.**
- **information is obtained from a non-representative sample of the population**
- **Sampling error can not be computed**
- **Survey results cannot be projected to the population**

3. Selecting a Sampling Design
Non-Probability Sampling

Advantages

- **Cheaper and faster than probability**
- **Reasonably representative if collected in a thorough manner**

Types of Non-Probability Sampling (I)

- ***Convenience Sampling***
  - **A researcher's convenience forms the basis for selecting a sample.**
    - **people in my classes**
    - **Mall intercepts**
    - **People with some specific characteristic (e.g. bald)**
- ***Judgement Sampling***
  - **A researcher exerts some effort in selecting a sample that seems to be most appropriate for the study.**

Types of Non-Probability Sampling

- ***Snowball Sampling***
  - **Selection of additional respondents is based on referrals from the initial respondents.**
    - **friends of friends**
  - **Used to sample from low incidence or rare populations.**
- ***Quota Sampling***
  - **The population is divided into cells on the basis of relevant control characteristics.**
  - **A quota of sample units is established for each cell.**
    - **50 women, 50 men**
  - **A convenience sample is drawn for each cell until the quota is met.**

  (similar to stratified sampling)

Probability Vs. Non-Probability Sampling

- **Non-probability sampling is less time consuming and less expensive.**
- **The probability of selecting one element over another is not known and therefore the estimates cannot be projected to the population with any specified level of confidence. Quantitative generalizations about population can only be done under probability sampling.**
- **However, in practice, marketing researchers also apply statistics to study non-probability samples.**

Errors in Survey

Random Sampling Error

- random error- the sample selected is not representative of the population due to chance
- the level of it is controlled by sample size
- a larger sample size leads to a smaller sampling error.

Population mean (μ) gross income = $42,300

Sample 1 (400/250,000) mean $(\overline{X})$ = $41,100

Sample 1 (400/250,000) mean $(\overline{X})$ = $43,400

Non-Sampling Errors (I)

Non-sampling Error

–systematic Error

–the level of it is NOT controlled by sample size.

- The basic types of non-sampling error
  - Non-response error
  - Response or data error
- A *non-response error* occurs when units selected as part of the sampling procedure do not respond in whole or in part
  - If non-respondents are not different from those that did respond, there is no non-response error

Non-Sampling Errors (II)

- A *response or data error* is any systematic bias that occurs during data collection, analysis or interpretation
  - Respondent error (e.g., lying, forgetting, etc.)
  - Interviewer bias
  - Recording errors
  - Poorly designed questionnaires