# 6
# Sampling

Research Methodology
2009

Monica A. Kappiantari

# 6 Sampling

**References:**

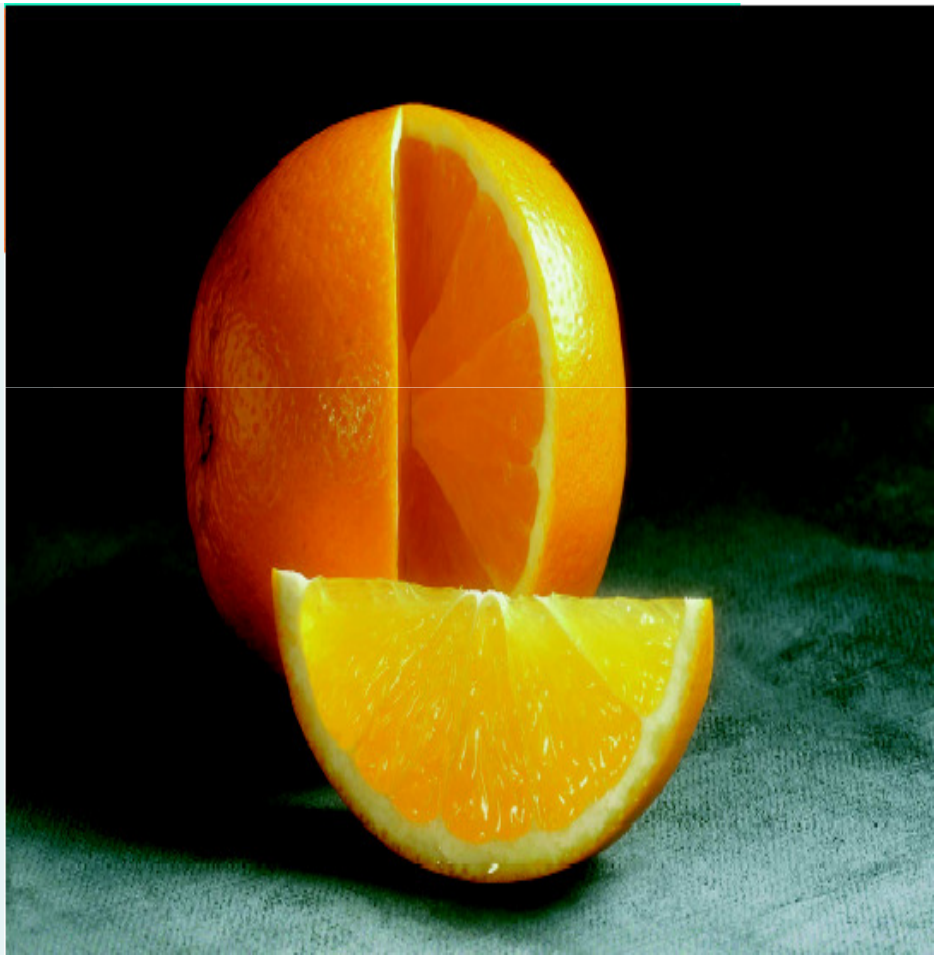
1. Cooper, Donal R., Pamela S. Schindler, *Business Research Methods,* 9th ed, Mc-Graw Hill International Edition, 2006, ch. 15.
2. http://business.nmsu.edu/~mhyman/

**Topics:**

1. Nature of Sampling
   1. Why sample
   2. Good sample
   3. Types of sample design
2. Steps in Sampling Design
3. Probability Sampling
4. Nonprobability Sampling

# The Nature of Sampling

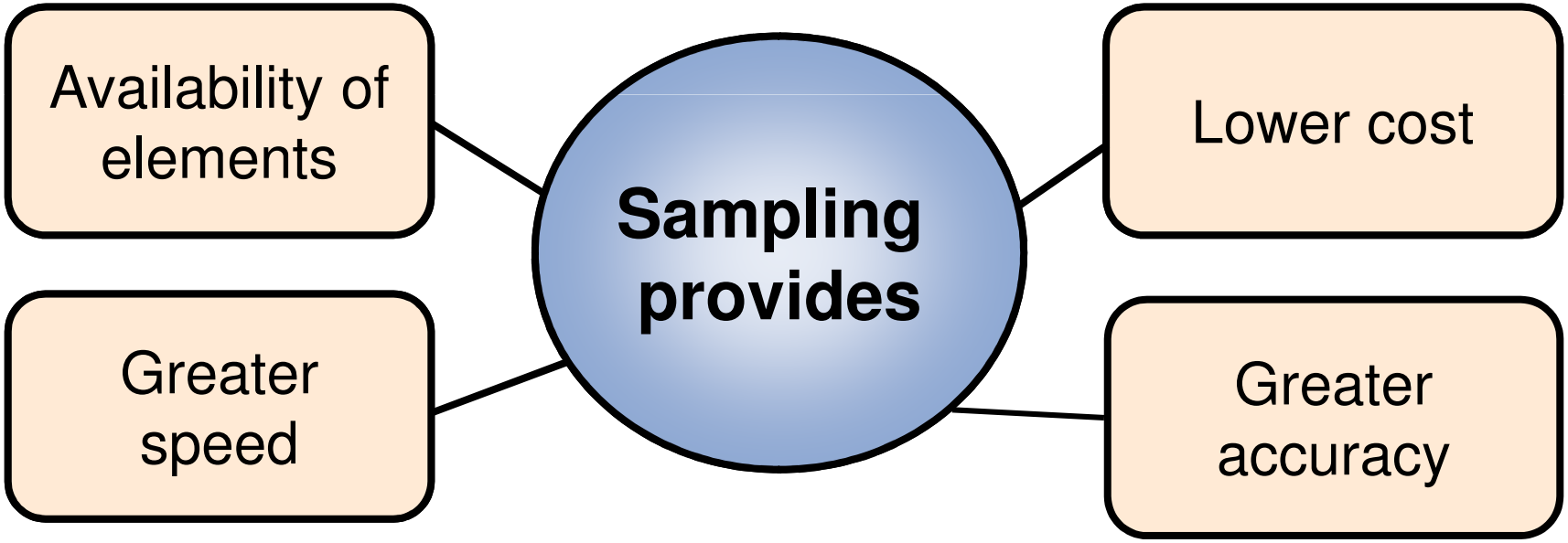
- Sampling
- Population Element
- Population
- Census
- Sampling frame



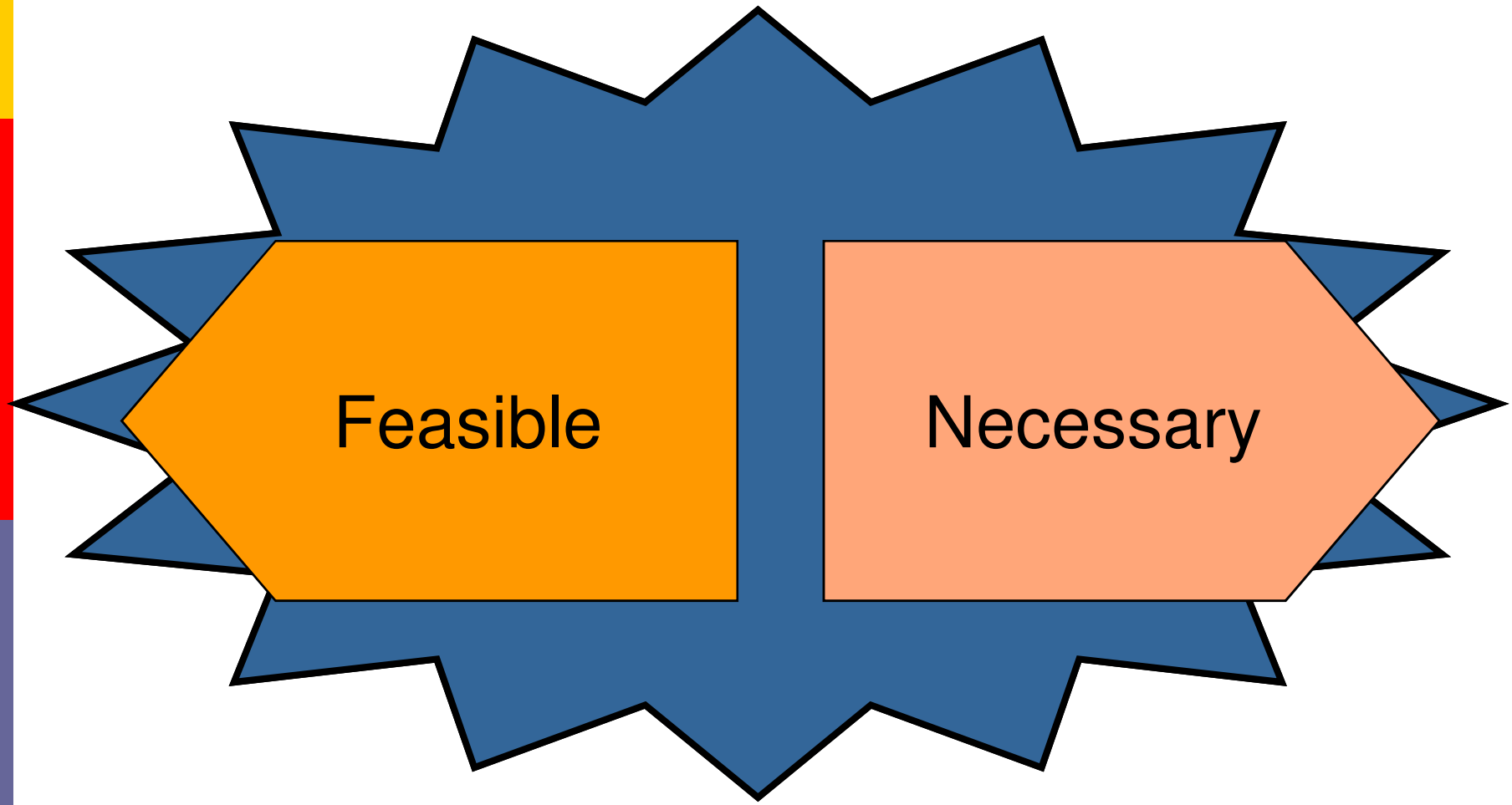
# The Nature of Sampling

- The **basic idea of sampling** is that by selecting some of the elements in a population, we may draw conclusions about the entire population.
- A **population element** is the individual participant or object on which the measurement is taken. It is the unit of study. It may be a person but it could also be any object of interest.
- A **population** is the total collection of elements about which we wish to make some inferences.
- A **census** is a count of all the elements in a population.
- The listing of all population elements from which the sample will be drawn is called the **sampling frame.**

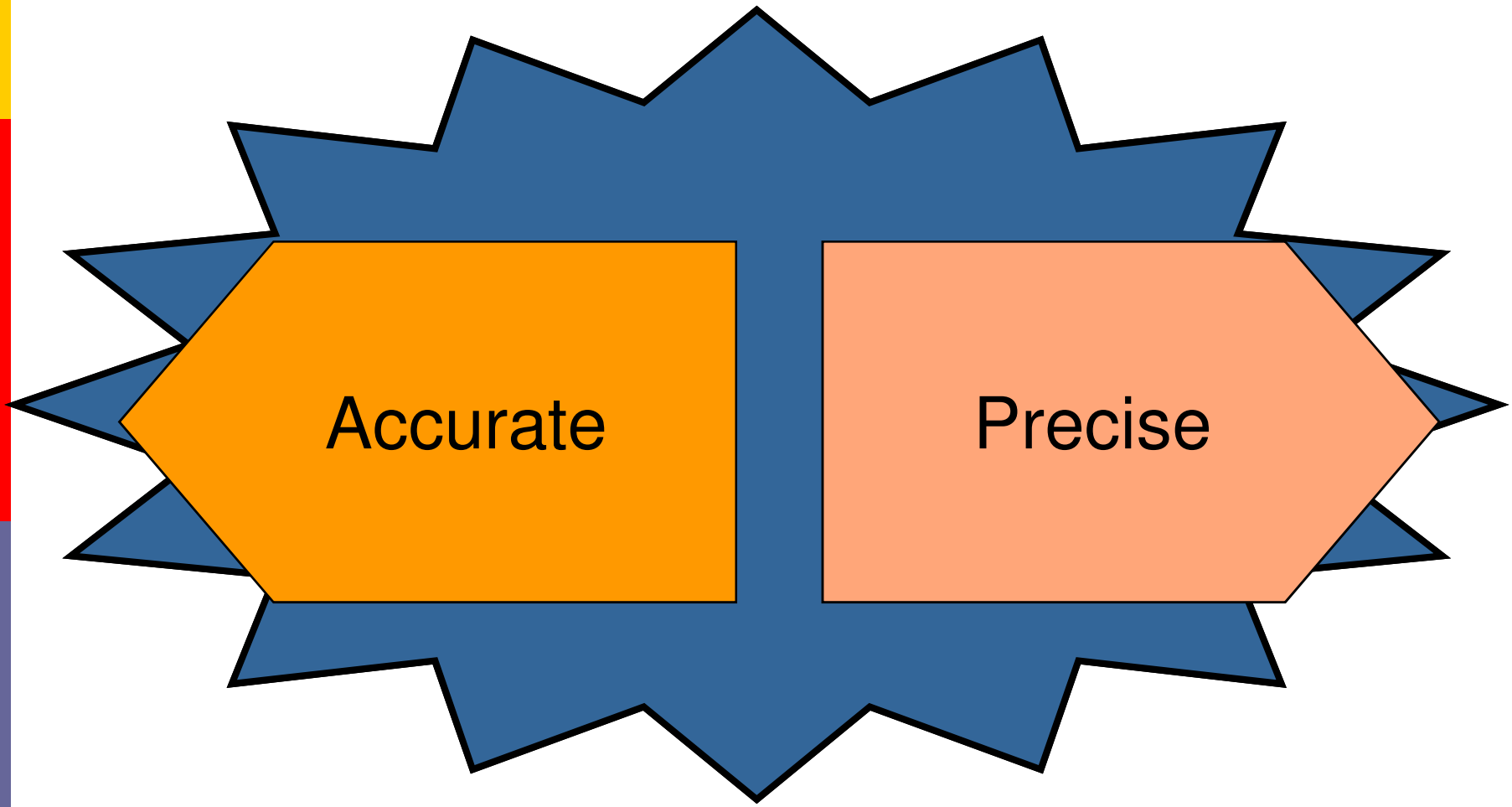
# Why Sample?


Availability of elements
Greater speed
Sampling provides
Lower cost
Greater accuracy

# When Is A Census Appropriate?



# What Is A Good Sample?



# What is a Good Sample?

- Accuracy is the degree to which **bias** is absent from the sample.
- When the sample is drawn properly, the measure of behavior, attitudes, or knowledge of some sample elements **will be less** than the measure of those same variables drawn from the population.
- Increasing the **sample size** can reduce systematic variance as a cause of error.
- **Systematic variance** is a variation that causes measurements to skew in one direction or another.

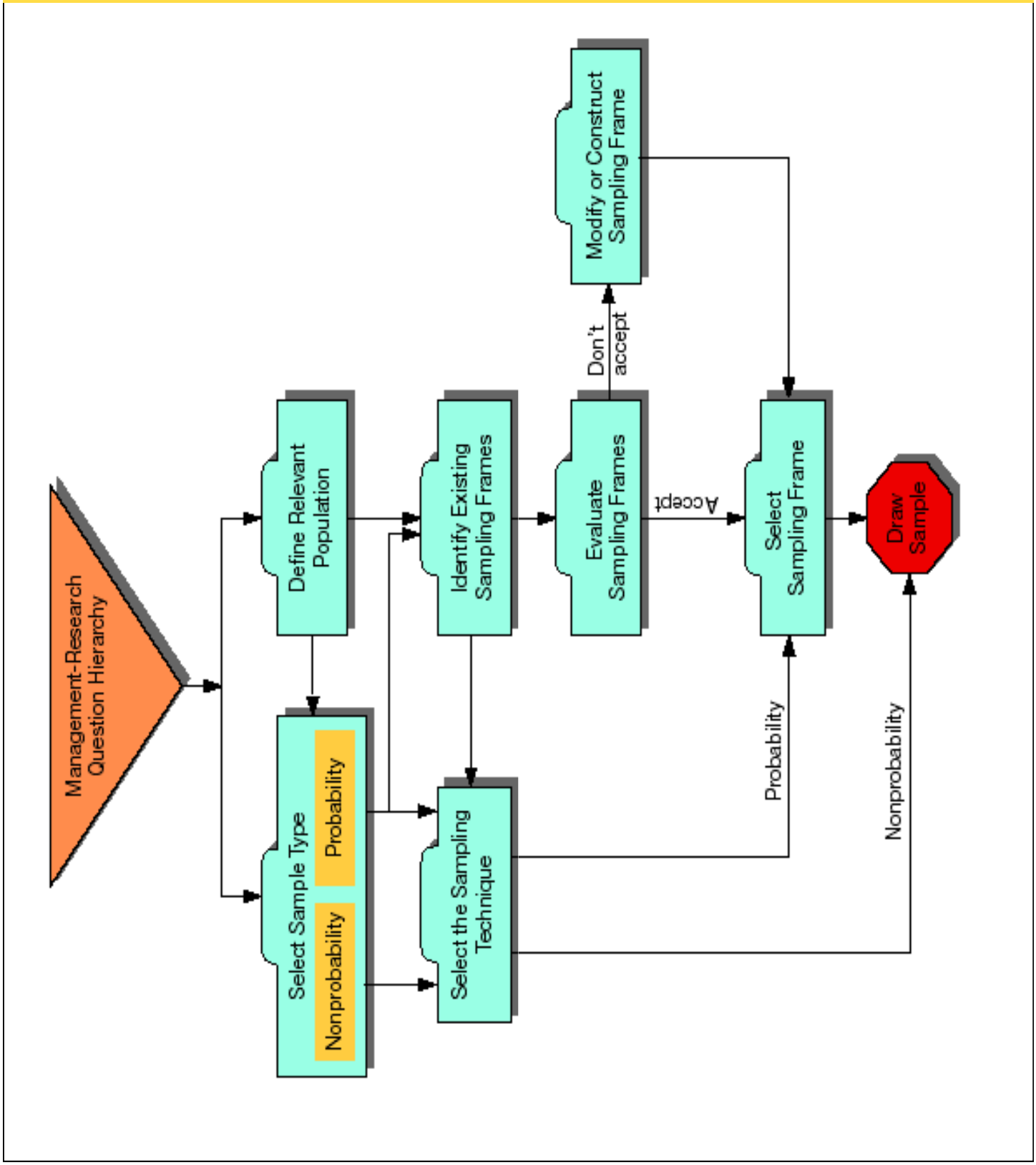
# What is a Good Sample?

- The numerical descriptors that describe samples may be expected to differ from those that describe populations because of random fluctuations inherent in the sampling process.
- This is called **sampling error** and reflects the influence of chance in drawing the sample members.
- Sampling error is what is left after all known sources of systematic variance have been accounted for.

# What is a Good Sample?

- **Precision** is measured by the **standard error of estimate**, a type of standard deviation measurement.
- The smaller the standard error of the estimate, the higher is the precision of the sample.

# Sampling Design within the Research Process



# Types of Sampling Designs

| Element Selection | Probability | Nonprobability |
|---|---|---|
| Unrestricted | Simple random | Convenience |
| Restricted | Complex random | Purposive |
| | Systematic | Judgment |
| | Cluster | Quota |
| | Stratified | Snowball |
| | Double | |

# Types of Sampling Design

- **Nonprobability sampling** is an arbitrary and subjective sampling procedure where each population element does not have a known, nonzero chance of being included.
- **Probability sampling** is a controlled, randomized procedure that assures that each population element is given a known, nonzero chance of selection.

# Steps in Sampling Design

What is the target population?

What are the parameters of interest?

What is the sampling frame?

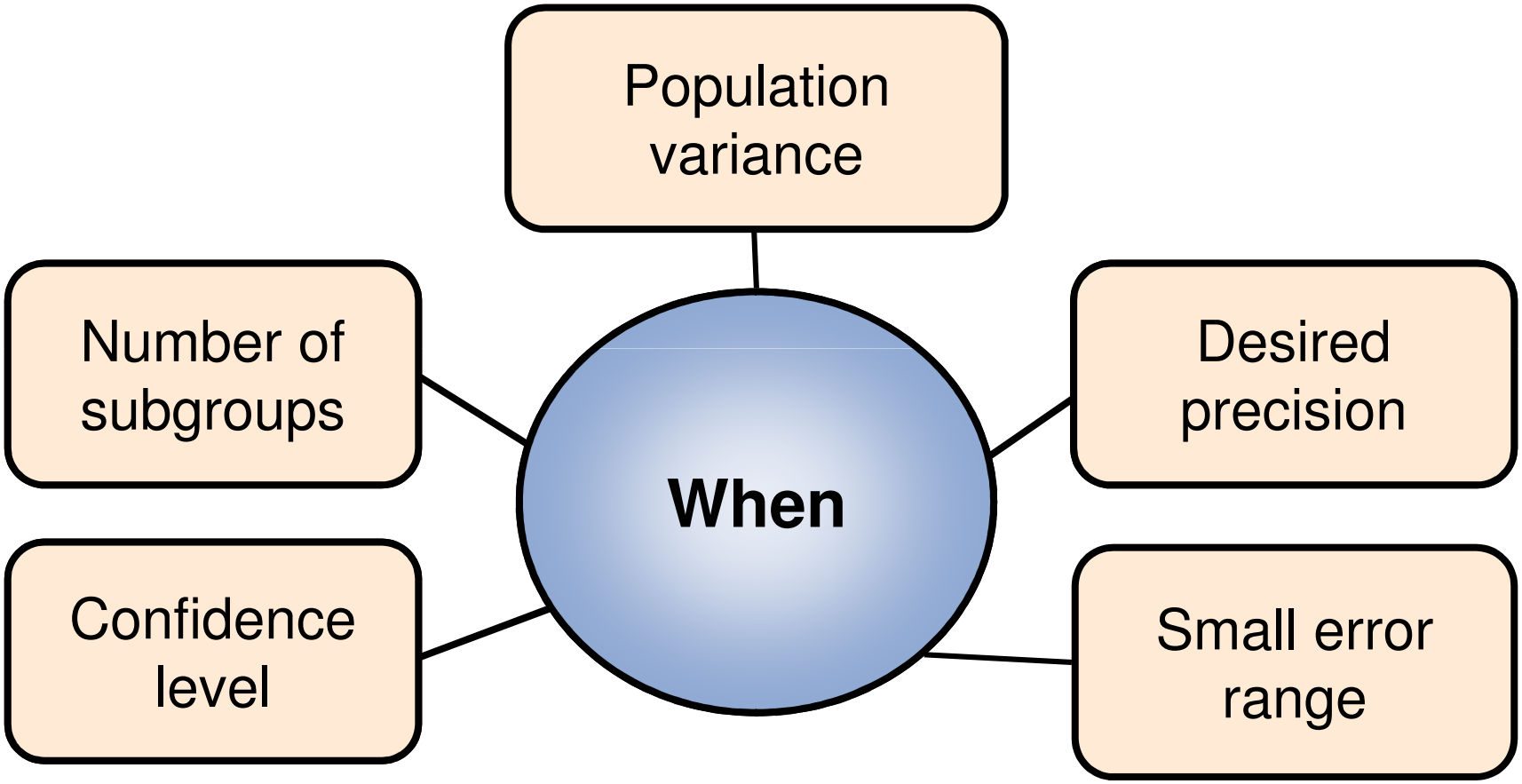
What is the appropriate sampling method?

What size sample is needed?

# Probability Sampling

- Simple Random Sampling
- Complex Probability Sampling
  - Systematic Sampling
  - Stratified Sampling
    - Proportionate
    - Disproportionate
  - Cluster Sampling
    - Area Sampling
    - Design
  - Double Sampling

# Larger Sample Sizes



# Simple Random

- Each population element has an **equal chance** of being selected into the samples.
- The sample is drawn using a random number table or **generator**.
- The **probability of selection** is equal to the sample size divided by the population size.

# Simple Random

The steps are as follows:

1) Assign each element within the sampling frame a unique number.
2) Identify a random start from the random number table.
3) Determine how the digits in the random number table will be assigned to the sampling frame.
4) Select the sample elements from the sampling frame.

# Simple Random

## Advantages

- Easy to implement with RDD (random digit dialing)

## Disadvantages

- Requires list of population elements
- Time consuming
- Uses larger sample sizes
- Produces larger errors
- High cost

# Systematic

- An element of the population is selected at the beginning with a **random start** and then every *k*th element is selected until the appropriate size is selected.
- The *k*th element is the **skip interval**, the interval between sample elements drawn from a sample frame in systematic sampling.
- It is determined by dividing the population size by the sample size.

# Systematic

To draw a systematic sample, the steps are:

1. Identify, list, and number the elements in the population
1) Identify the skip interval
2) Identify the random start
3) Draw a sample by choosing every *k*th entry.

To protect against subtle biases, the research can

1) Randomize the population before sampling,
2) Change the random start several times in the process
3) Replicate a selection of different samples.

# Systematic

## Advantages

- Simple to design
- Easier than simple random
- Easy to determine sampling distribution of mean or proportion

## Disadvantages

- Periodicity within population may skew sample and results
- Trends in list may bias results
- Moderate cost

# Stratified

- The population is divided into **subpopulations** or strata and uses simple random on each strata.
- Results may be weighted or combined.
- The cost is high.
- Stratified sampling may be proportion or disproportionate.
- In **proportionate stratified sampling**, each stratum's size is proportionate to the stratum's share of the population.
- Any stratification that departs from the proportionate relationship is **disproportionate**.

# Stratified

**Advantages**

- Control of sample size in strata
- Increased statistical efficiency
- Provides data to represent and analyze subgroups
- Enables use of different methods in strata

**Disadvantages**

- Increased error will result if subgroups are selected at different rates
- Especially expensive if strata on population must be created
- High cost

# Cluster

- The population is divided into internally heterogeneous subgroups, some are randomly selected for further study.
- Two conditions foster the use of cluster sampling:
  - The need for **more economic efficiency** than can be provided by simple random sampling
  - the frequent **unavailability of a practical sampling frame** for individual elements.

# Cluster

Several questions must be answered when designing cluster samples:

1) How homogeneous are the resulting clusters?
2) Shall we seek equal-sized or unequal-sized clusters?
3) How large a cluster shall we take?
4) Shall we use a single-stage or multistage cluster?
5) How large a sample is needed?

# Cluster

**Advantages**

- Provides an unbiased estimate of population parameters if properly done
- Economically more efficient than simple random
- Lowest cost per sample
- Easy to do without list

**Disadvantages**

- Often lower statistical efficiency due to subgroups being homogeneous rather than heterogeneous
- Moderate cost

# Stratified and Cluster Sampling

## Stratified

- Population divided into few subgroups
- Homogeneity within subgroups
- Heterogeneity between subgroups
- Choice of elements from within each subgroup

## Cluster

- Population divided into many subgroups
- Heterogeneity within subgroups
- Homogeneity between subgroups
- Random choice of subgroups

# Area Sampling

Area sampling is a cluster sampling technique applied to a population with well-defined political or geographic boundaries. It is a low-cost and frequently used method.

# Double

- In drawing a sample with double (sequential or multiphase) sampling, data are collected using a previously defined technique.
- Based on the information found, a **subsample** is selected for further study.

# Double

## Advantages

- May reduce costs if first stage results in enough data to stratify or cluster the population

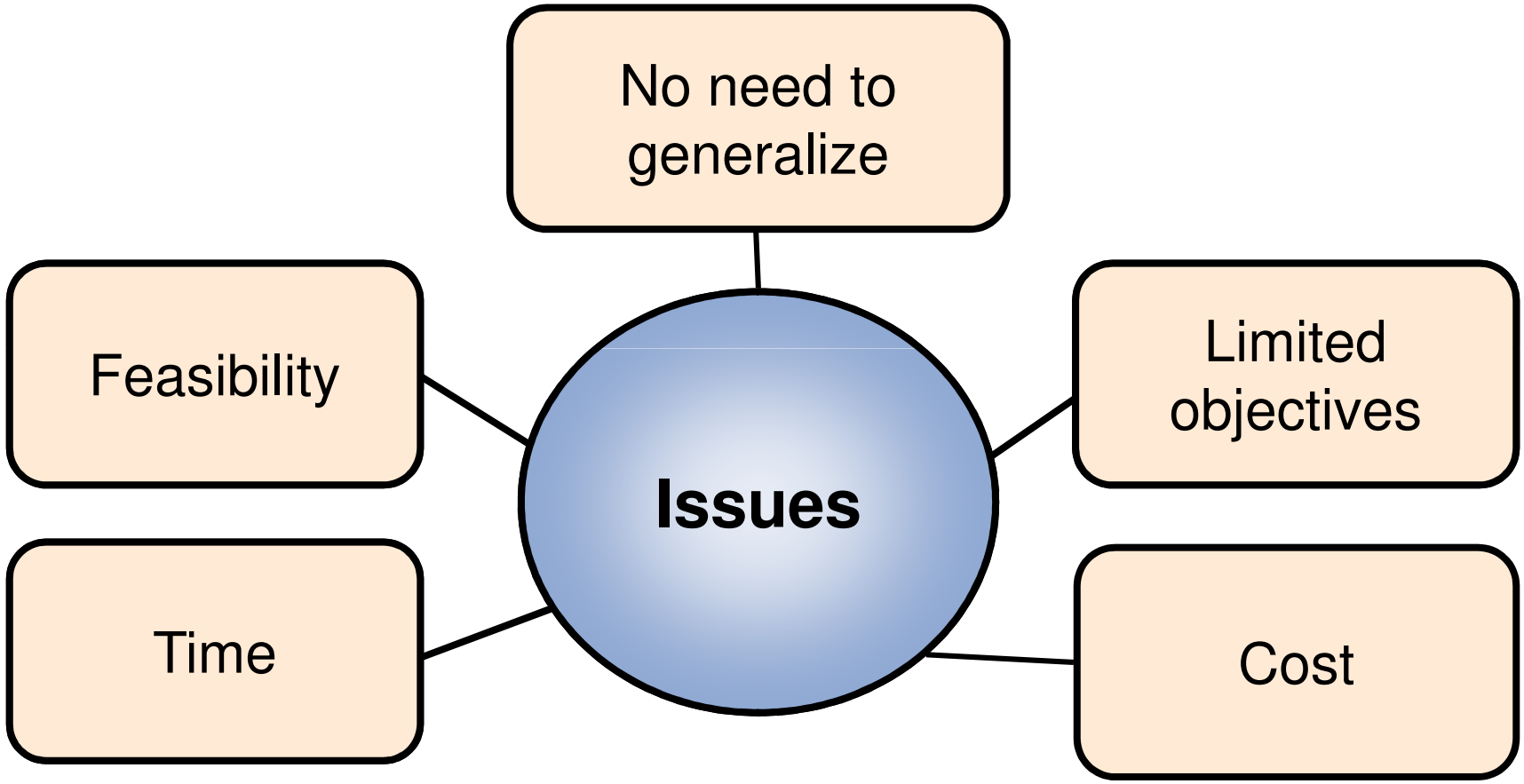
## Disadvantages

- Increased costs if discriminately used

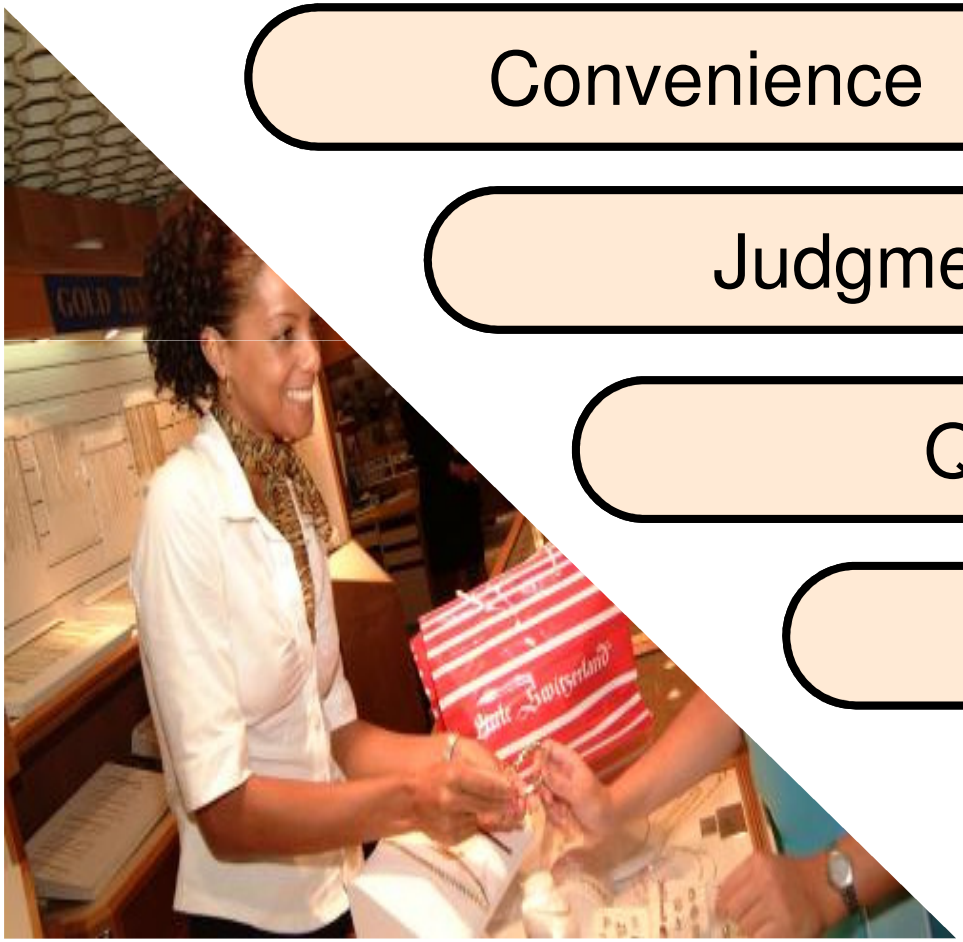
# Nonprobability Sampling

- Convenience
- Purposive
  - Judgement
  - Quota
- Snowball

# Nonprobability Samples



# Nonprobability Sampling Methods



- Convenience
- Judgment
- Quota
- Snowball

# Convenience

- Convenience samples are nonprobability samples where the element selection is based on **ease of accessibility**.
- They are the least reliable but **cheapest** and **easiest** to conduct.
- Examples include informal pools of friends and neighbors, people responding to an advertised invitation, and "on the street" interviews.

# Judgement

- Judgment sampling is purposive sampling where the researcher **arbitrarily selects sample units** to conform to some criterion.
- This is appropriate for the **early stages of an exploratory study**.

# Quota

- Relevant characteristics are used to **stratify the sample** which should improve its representativeness.
- The logic behind quota sampling is that certain relevant characteristics describe the dimensions of the population.
- In most quota samples, researchers specify more than one control dimension.
- Each dimension should have a distribution in the population that can be estimated and be pertinent to the topic studied.

# Snowball

- Snowball sampling means that subsequent participants are referred by the current sample elements.
- This is useful when **respondents are difficult to identify** and best located through referral networks.
- It is also used frequently in **qualitative studies.**