Chapter 1
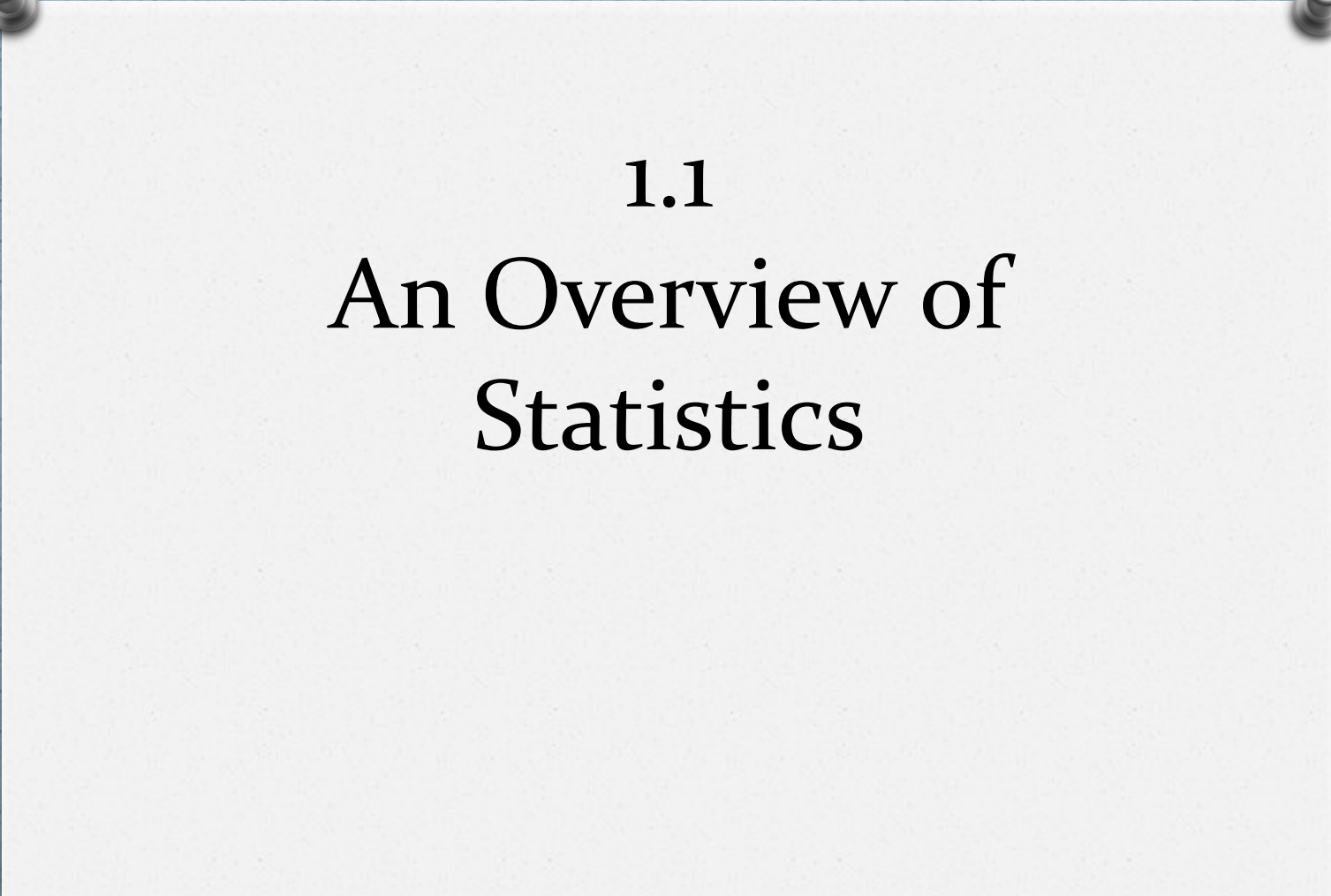# Introduction to Statistics

1.1
An Overview of Statistics

Definitions:

O Data – _____

   O **Population** – _____ being studied

   O **Sample** – _____of the population

O Statistics – the science of collecting, organizing, analyzing, and interpreting _____ in order to _____ _____.

EX: Identify the population and the sample

O In a recent survey, 614 small business owners in the U.S. were asked whether they thought their companies Facebook presence was valuable.

O The U.S. Department of Energy conducts weekly surveys of approximately 800 gas stations to determine the average price per gallon of regular gas.

Definitions:

O **Parameter** – a numerical description of a

   _____

O **Statistic** – a numerical description of a

   _____

EX: Determine whether the numerical value is a parameter or a statistic.

O A recent survey of approximately 400,000 employers reported that the average starting salary for marketing majors is $53,000.

O The freshman class at a university has an average SAT math score of 514.

O In a random check of 400 retail stores, the FDA found that 34% of the stores were not storing fish at the proper temperature.
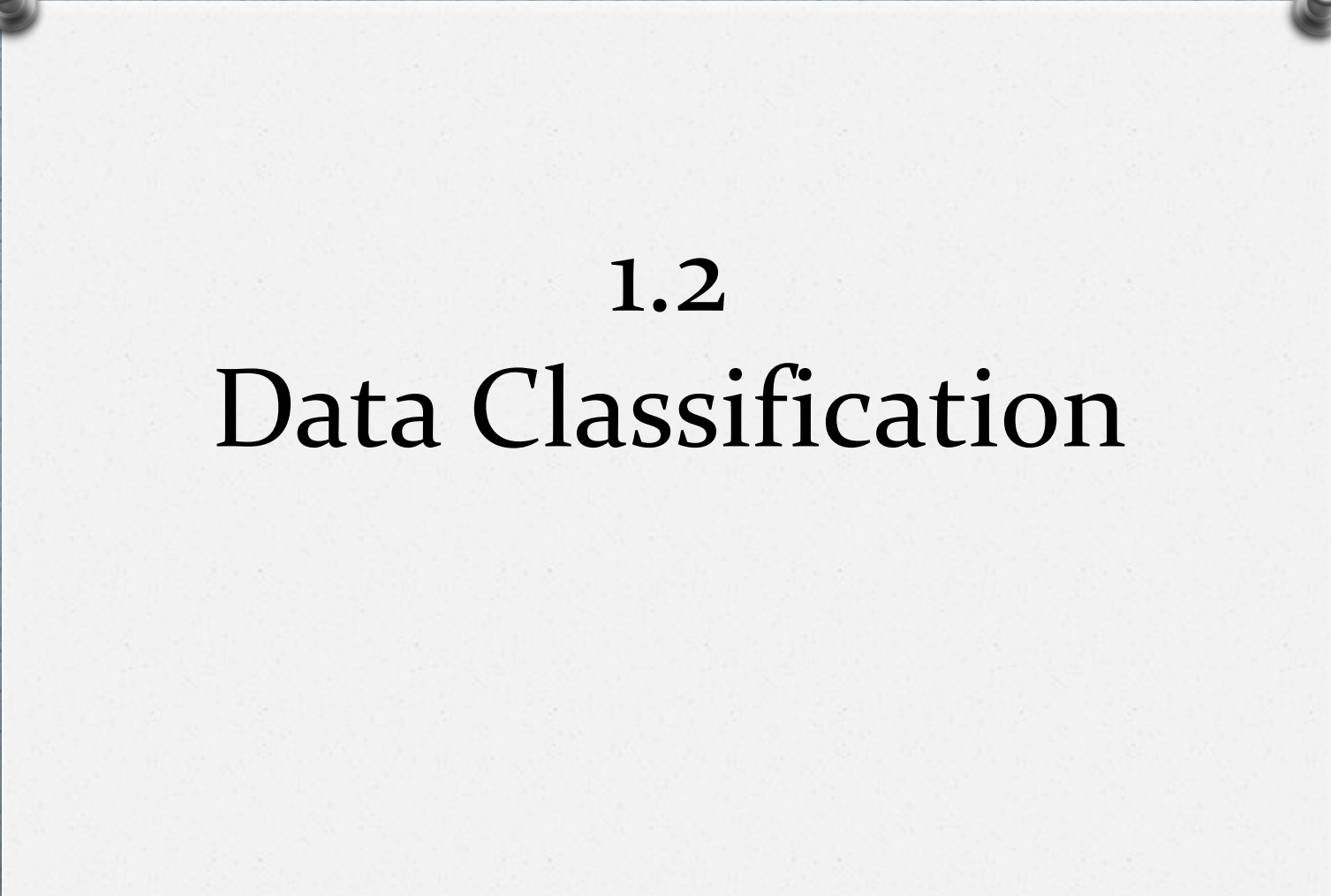
Branches of Statistics

O **Descriptive Statistics** – the branch of stats that involves _____

_____

O **Inferential Statistics** – the branch of stats that involve using a sample to _____

_____ about a population

EX:

O A large sample of men, aged 48, was studied for 18 years. For unmarried men, approximately 70% were alive at age 65. For married men, 90% were alive at age 65.

O A) Which part of the study represents descriptive statistics?

O

B) What conclusions might be drawn from the study using inferential statistics?

1.2
Data Classification

Types of Data

O Qualitative Data – consists of _____
_____

    O EX:

O Quantitative Data – consists of _____
_____

    O EX:

EX:

Name	StartYear	EndYear	Position	LEN	Height	Weight	BMI
Tiny Archibald*	1971	1984	G	13	6-1	150	19.78795
Chet Aubuchon	1947	1947	G	0	5-10	137	19.65531
Mike Barrett	1970	1973	G	3	6-2	155	19.89865
Ron Behagen	1974	1980	C-F	6	6-9	185	19.82244
Manute Bol	1986	1995	C	9	7-7	200	16.97863
Joe Bryant	1976	1983	C-F	7	6-9	185	19.82244
Barney Cable	1959	1964	F	5	6-7	175	19.71239
Keith Closs	1998	2000	C	2	7-3	212	19.69032
Trey Gilder	2010	2010	F	0	6-9	185	19.82244
Thomas Hamilton	1996	2000	C	4	7-2	330	31.36696
Mickey Johnson	1975	1986	F	11	6-10	190	19.86466
Bill Jones	1989	1989	F	0	6-7	175	19.71239
Oliver Miller	1993	2004	C	11	6-9	280	30.00152
Boniface N'Dong	2006	2006	C	0	7-0	198	19.72704
Craig Neal	1989	1991	G	2	6-5	165	19.56401
Chuck Nevitt	1983	1994	C	11	7-5	217	19.25906
Shaquille O'Neal	1993	2011	C	18	7-1	325	31.62284
Louis Orr	1981	1988	F	7	6-8	175	19.22266
Red Rocha	1948	1957	C-F	9	6-9	185	19.82244
Garret Siler	2011	2011	C	0	6-11	305	31.12426
Mike Sweetney	2004	2007	F	3	6-8	275	30.20703
Robert Traylor	1999	2005	F	6	6-8	284	31.19562
Jahidi White	1999	2005	C-F	6	6-9	290	31.07301

EX: Determine whether the data is qualitative or quantitative.

O Jersey numbers of soccer players

O Mile times of runners

O Hair color of your classmates

Levels of Measurement

O Nominal – _____ data only

   O No _____ can be done with data

   O EX:

O Ordinal – _____ or _____ data

   O Data can be _____ but _____ are meaningless

   O EX:

O Interval – _____ data only

   O Data can be _____, differences
     are _____, but there is no

     _____

     _____

   O EX:

O Ratio – _____ data only

   O Data can be _____, differences are _____, there is a _____, and _____ are meaningful

   O EX:
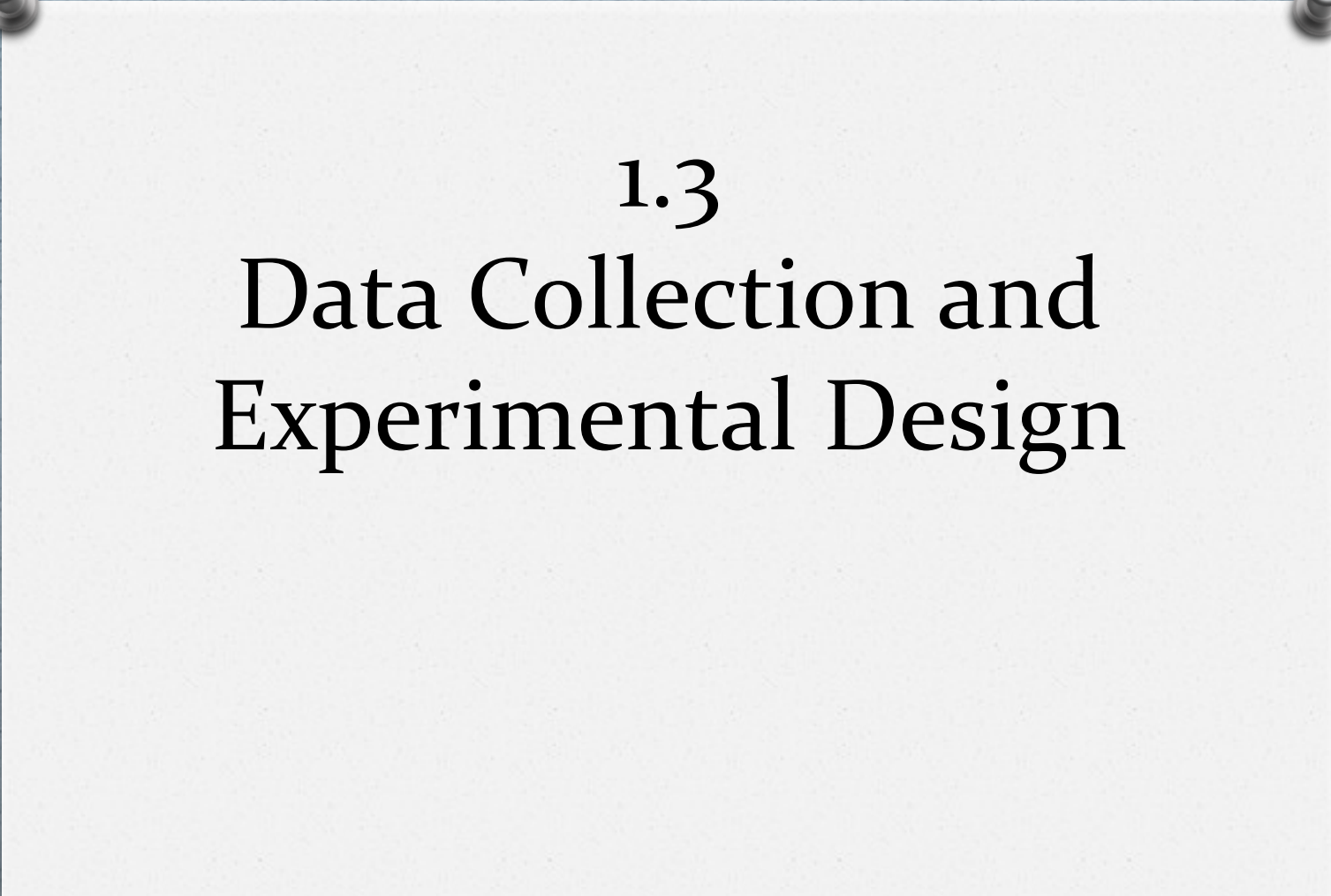
EX: Determine the level of measurement.

O The final standings for the Pacific Division of the NBA

O A collection of phone numbers

O The hourly body temperatures of a preemie in the NICU

O The heart rates (in beats per minute) of a preemie in the NICU

1.3
Data Collection and Experimental Design

Statistical Studies

O Goal: To _____ data and then _____the data to make a _____ about the whole population.

O Note: Any decision that is made using the results is only as good as the process used to obtain the data

    O If the process is _____the results will be _____

Types of Studies

O Observational – a researcher does _____ _____ the responses

    O Just _____ are made

    O _____ are made to existing conditions

Types of Studies

O Experiment – a researcher deliberately _____ before observing the responses

- O Treatment Group – part of the population that _____ the treatment

- O Control Group – part of the population that does _____ the treatment

- O Experimental Units – _____ in both groups

- O Placebo – a _____ treatment, that is made to look like the _____ treatment, often given to the _____

O Placebo effect – when a subject _____ _____ to a placebo when in fact the subject was given a _____ _____.

EX: Determine whether the study is an observational study or an experiment.

O Researchers study the effect of vitamin D supplementation among patients with antibody deficiency or frequent respiratory tract infections. To perform the study, 70 patients receive 4000 IU of vitamin D daily for a year. Another group of 70 patients receive a placebo daily for one year.

O Researchers conduct a study to find the U.S. public approval rating of the U.S. president. To perform the study, researchers call 1500 U.S. residents and ask them whether they approve or disapprove of the job being done by the president.

Data Collection

O Simulation – the use of a _____ to _____ of a situation or process.

  O EX:

O Survey – an _____ of one or more characteristics of a population

  O EX:

  O Make sure questions are worded in a way that does not lead to _____ results

Experimental Design

O To produce meaningful and unbiased results, experiments should be carefully designed and executed.

O Elements of a well-designed experiment:

O 1)

O 2)

O 3)

1) Control

O _____ must be controlled:

O Confounding variable – occurs when an experimenter cannot tell the difference between different factors on the variable

O EX:

2) Randomization

O A process of _____assigning subjects to _____treatment groups

   O Completely randomized design – treatment groups chosen by _____

3) Replication

O The _____ of an experiment under the _____ conditions.

O Sample size – the _____ of subjects in a study

O EX:

EX:

O A company wants to test the effectiveness of a new gum developed to help people quit smoking. Identify a potential problem with the given experimental design and suggest a way to improve it:

   O A company identifies ten adults who are heavy smokers. Five are given the new gum and the other five are given a placebo. After two months, the subjects are evaluated and it is found that the five subjects using the new gum have quit smoking.

O The company identifies 1,000 adults who are heavy smokers. The subjects are divided into blocks according to gender. Females are given the new gum and males are given the placebo. After two months, a significant number of the female subjects have quit smoking.

O The company identifies 240 adults who are heavy smokers. The subjects are randomly assigned to be in a treatment group or in a control group. Each subject is also given a DVD featuring the dangers of smoking. After 4 months, most of the subjects in the treatment group have quit smoking.

Sampling Techniques

O Census – a count or measure of the

_____

O Sampling – a count or measure of

_____

    O Must be _____ of the entire population

    O If it is not representative of the population, a_____ has occurred.

Sampling Techniques

O **Random sample – every member of the _____ has an _____ of being selected**

O Simple random sample – every possible _____has the same _____ of being selected

O To collect a SRS:

O Assign a _____ to each member of the population

O Then use a calculator or computer program to

_____

EX:

O There are 731 students enrolled in a stats course at your school. You wish to form a sample of 8 students to answer some survey questions. Select the students who will belong in the SRS using your calculator.
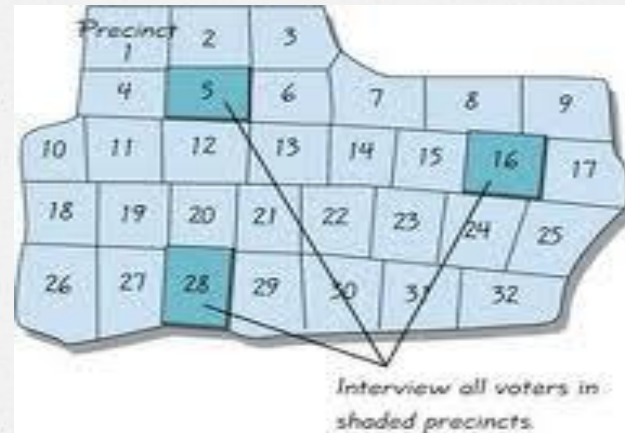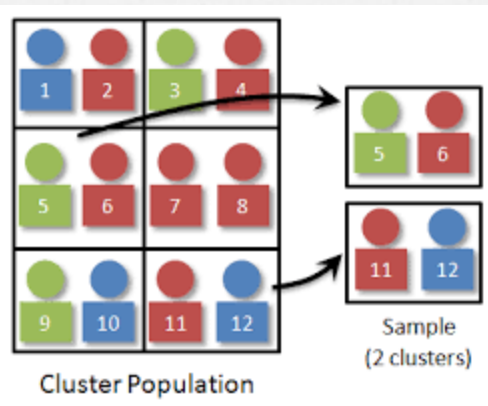
Other sampling techniques:

O Stratified sample – population is _____ in to two or more _____, called _____, that share similar characteristics. A sample is then _____ selected from _____strata.

O EX:
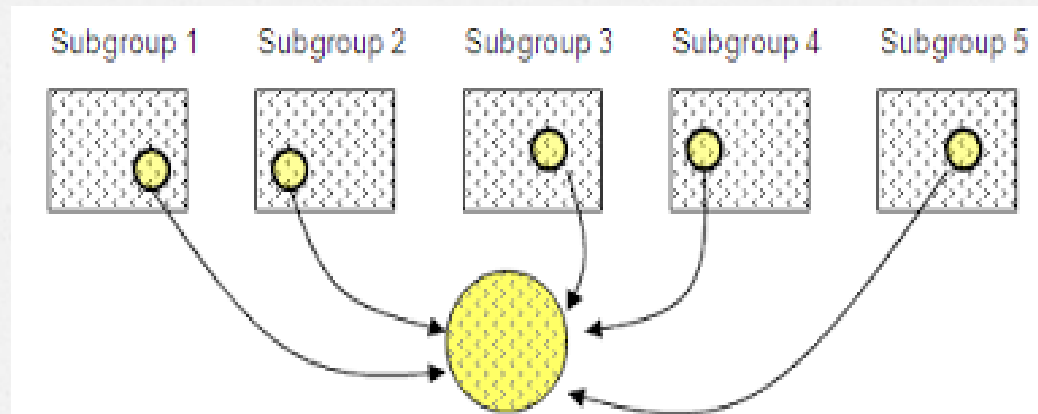
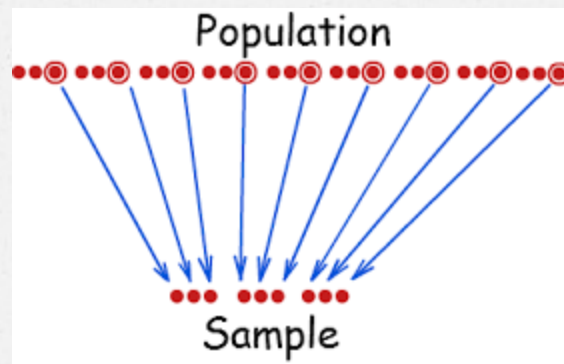O Cluster sample – population is divided into _____, called _____, and _____ in one or more clusters are selected

O EX:

Cluster Population
Sample (2 clusters)

Interview all voters in shaded precincts.

O Systematic sample – each member of the population is assigned a number. Members are _____ and selected at

_____
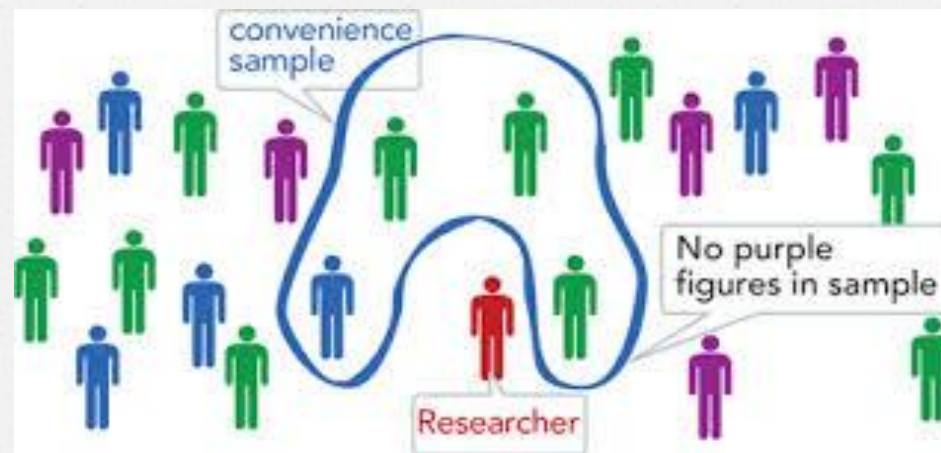
O EX: every 3rd person

O Convenience sample – consists of members of the population that are _____ to get

    O Often _____ so _____ recommended

    O EX: asking your friends around you

EX:

O You are doing a study to determine the opinions of students at your school regarding stem cell research. Identify the sampling technique used. Discuss any bias (if any).

 O You divide the student population with respect to majors and randomly select and question some students in each major.

O You assign each student a number and generate random numbers. You then question each student whose number is randomly selected.

O You select students who are in your biology class.

O You select a class at random and question each student in the class.

O You assign each student a number and, after choosing a starting number, question every 25th student.