Math-3

Lesson 7-2

Statistical Studies

Why Statistics?

You are sick!

Which is better, Brand ‘X’ or Brand ‘Y’ medicine in helping you get better from some disease?

How can we tell if the medicine actually works?

How can we tell the difference between the help the medicine is giving us and some other factor which might be helping us get better (a “confounding factor”)?

Research Question (or a “Statistical Question”) involves measurements that has some variability in the quantity being measured.

Requires us to take data then analyze the data to answer our question.

How tall are Americans?

How much does it rain in Utah?

Is Brand ‘X’ more effective than Brand “Y” in relieving pain?

Research Question

Which one is not a statistical question? If it is not, why not?

1. How many people will vote in the next Presidential election?

2. How long is the train?

No variability in the quantity being measured.

3. How much does Brand X cost?

Research Question

What portion of eligible voters will vote?

How could you find this out?

Method #1 Ask every eligible voter if he/she will vote.

Is this practical?

Too many eligible voters!

Take too long

Cost too much

Research Question

What portion (percentage) of eligible voters will vote?

How could you find this out?

Method #2: Take a sample.

Ask some (not all) of eligible voters if they will vote then hope that the percentage of those we asked who said they would vote is the same as the percentage of those who will actually vote (that we didn't ask).

How could we make it more likely that our result reflects the actual result for the entire eligible voter population?

Ask as many as we can.

Make sure our "sample" is NOT "biased".

Sampling a population: selecting certain members of the population and measuring some attribute of those members.

Sample size: that portion of the total population that was measured (not all) (equates to the number of measurements taken).

In method #2 (taking a sample) we only selected a certain portion of the total population. We measured this sample (we asked them whether they would vote or not).

What factors do you think could cause our result to NOT be the same as the rest of the population?

If we only ask those who intended to vote.

If we only ask those who intended to NOT vote.

How we decided who to ask.

The method used to get the data (example: online survey vs. mail-out survey).

Geographic area of those we ask.

Gender of those we ask.

“Socio-economic status” of those we ask.

More???

Sample Bias

What does sample bias mean?

Statistical sampling bias. a systematic (as opposed to a random distortion) of a statistic as a result of an error in the sampling procedure.

Example of sample bias: if we ask people in Utah which political party they are voting for, does it reflect the national voting trend?

Why not?

How do we minimize sample bias?

Make sure those we ask are RANDOMLY selected.

What does random mean?

Randomization

A sample population is considered *random* if the probability of selecting a particular sample is the same as the probability of selecting every other sample. When a sample is not random, a bias is introduced which may influence the study in favor of one outcome over other outcomes.

How do we ensure each person has an equal chance of being selected?

Phone books? Tax records? Birth certificates?

Social security numbers?

The design of the experiment is crucial in minimizing bias.

Summary (so far)

We need to know something about a population (set), we need a ***question that involves taking data*** (then evaluating the data) to answer the question.

The ***population may be too large to measure every member*** of the population.

We will ***sample a portion of the population*** to measure.

We ***design a method*** of selecting each data point so that the selection process *is totally random*.

IF the sample is randomly drawn, then the sample measurement fairly accurately represents the measurement we would get if we measured the entire population.

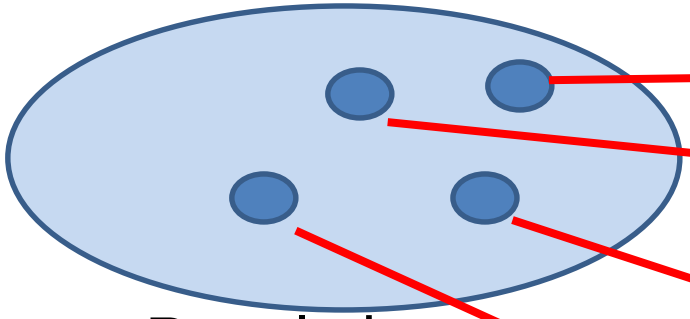
Types of Statistical Studies

1. Sample Study
2. Experimental Study
3. Observational Study

Equation for the graph of the bell curve (normal distribution).

$$p(x) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

1. <u>"Sample Study"</u> - The purpose of a sample study is to *<u>estimate</u>* a certain parameter of a population.

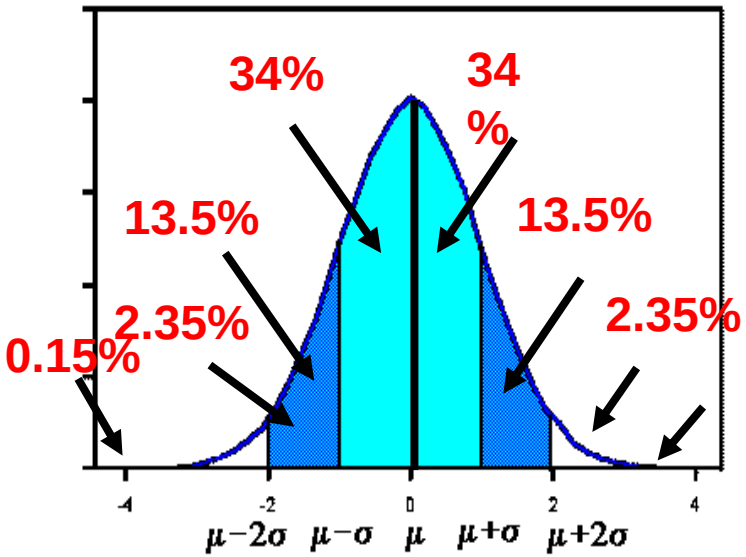


Large Population

<u>Random</u> Sample

105
100
100
100
95
90
85
80
80
75
75
75
75
75
70
65
55
50
50
50
45
35

Data set



Calculate some <u>statistics</u> that summarize the data.

1st type of statistical study

1. "Sample Study" - The purpose of a sample study is to *estimate* a certain parameter of a population.

Example: a **survey** is a sample study. It may focus on opinions or factual information depending upon the purpose of the study. Although a survey may ask more than one question, each question stands alone so the survey is estimating a certain parameter of the population.

Advantages of a Survey

- can be administered in a variety of forms (telephone, mail, on-line, mall interview, etc.)

- are efficient for collecting data from a large population

- are applicable to a wide range of topics

- can be designed to focus only on the needed response questions

Disadvantages of Surveys

- are dependent upon the respondent's honesty and motivation when answering

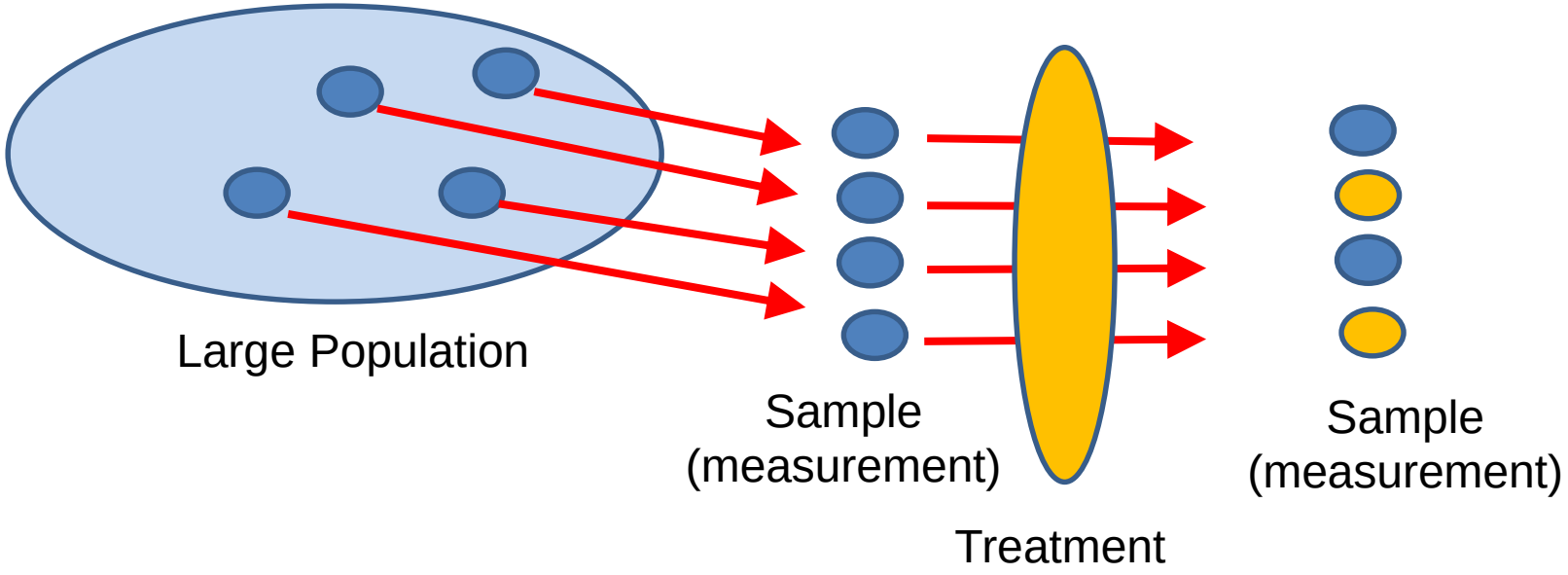
- can be flawed by non-response

- Sometimes participation in a survey, is “self-selecting” (people chose online surveys they feel strongly about).

- can possess questions or answer choices that may be interpreted differently by different respondents (such as the choice "agree slightly")

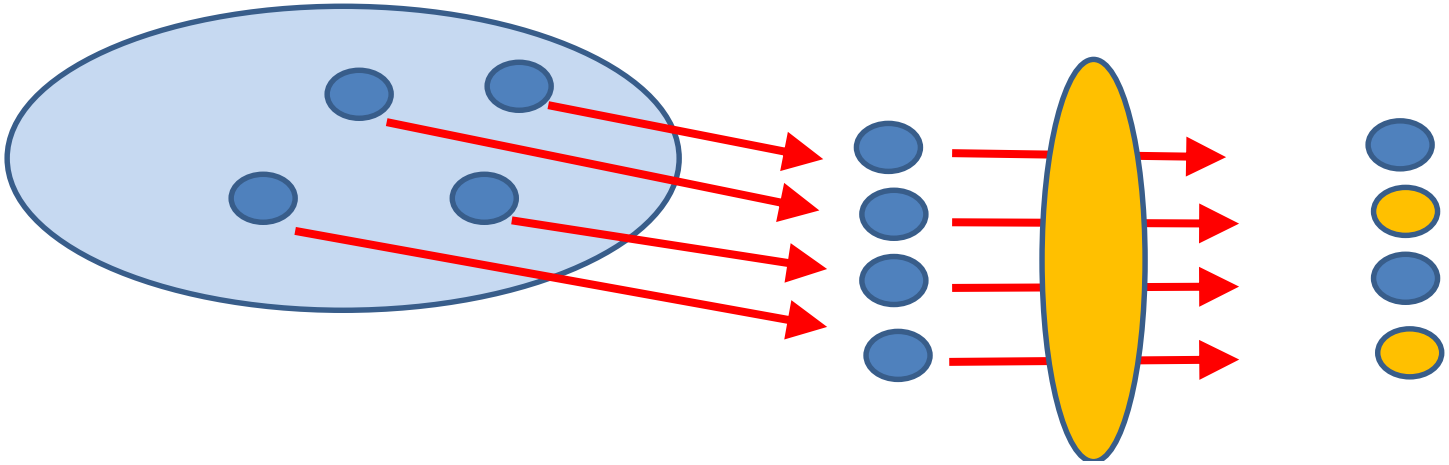
2nd Type of Statistical Study

2. **Experimental study**: a population is sampled. The sampled population is manipulated (fed a special diet, given a drug, etc.). The sampled population is "re-sampled" (re-measured) using the same measurement. The purpose is to see if the "manipulation" has some effect on the measurement.

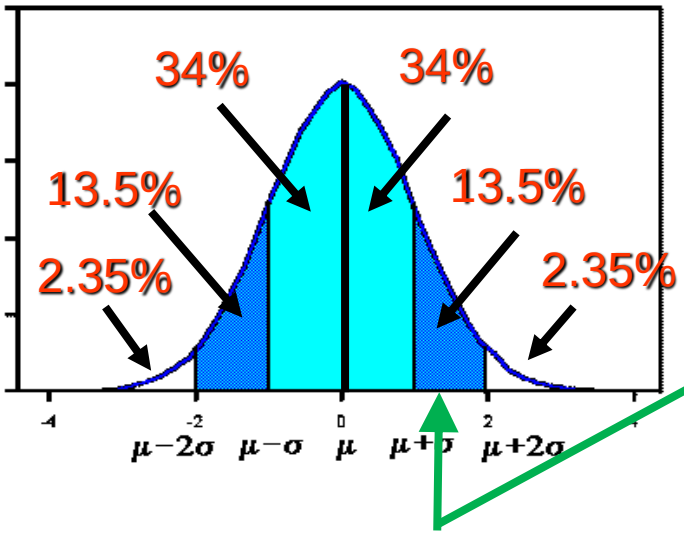


2. Experimental study:

Large Population



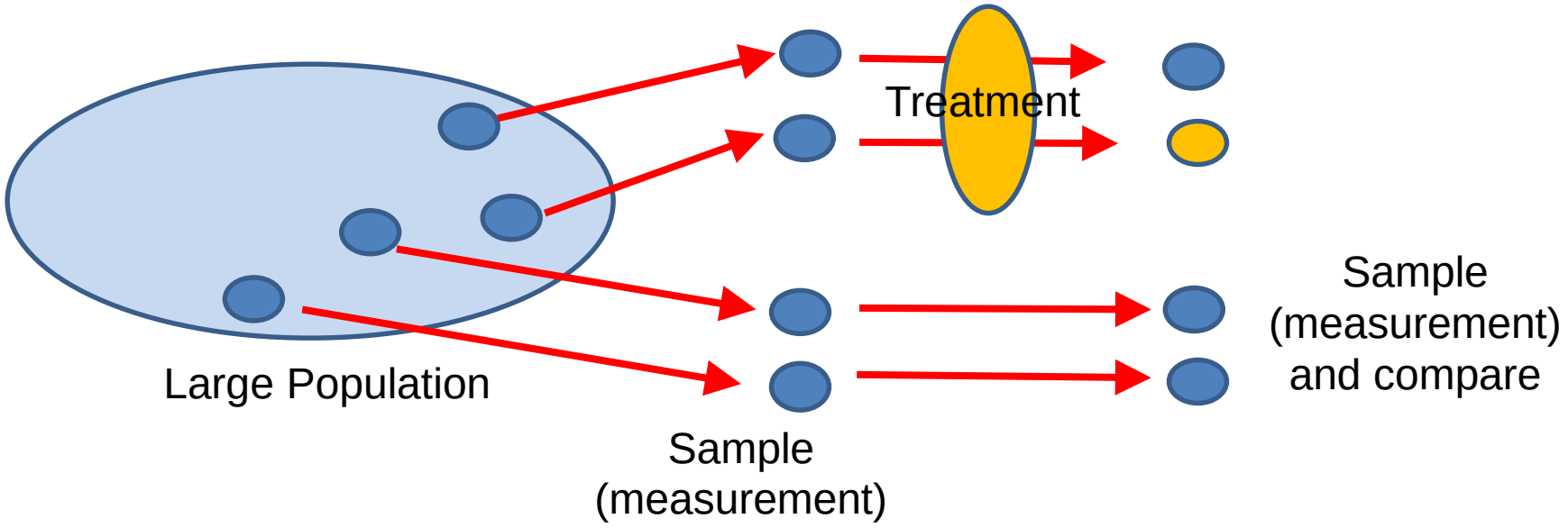
1. Sample (measure)



2. Post-treatment sample (measurement)

3. Compare statistics (mean of the treatment group) to the “expected” values (pre-treatment)

"Controlled" experimental study: the sample population is divided into groups with one group established as the control group. All groups (except for the control group) will be manipulated in some manner. Measurements are taken from each group and compared to the control group to see if the manipulation had an effect compared to the group that was not manipulated.



We compare the statistics of the two groups and use inference to predict what will happen in the general population.

Example of a controlled experimental study

1000 adults between the ages of 25 to 40 were randomly selected from across the country. All individuals were weighed.

Each name had a number assigned to it. 500 numbers were randomly selected.

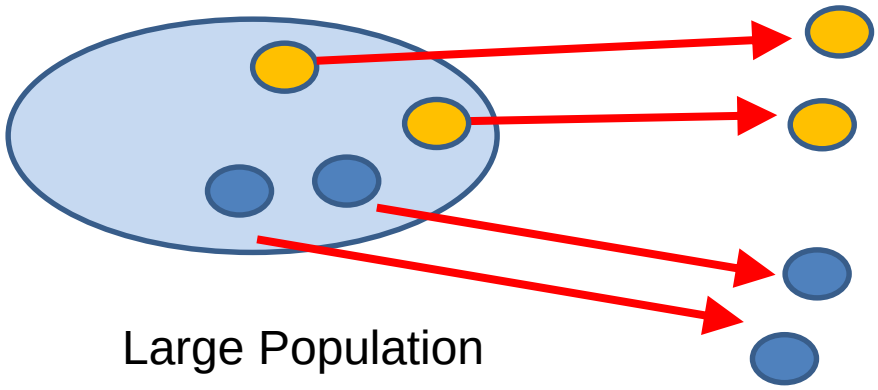
Odd numbers were assigned to the control group. A ***placebo*** is given to the control group.

Even numbers were assigned to the treatment group. These individuals were given a daily dose of Brand 'X'.

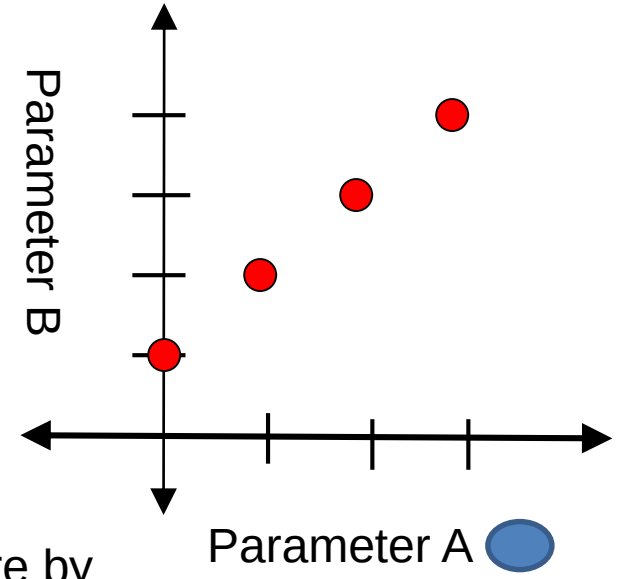
After 10 weeks of treatment, all individuals were weighed.

The mean weight of each group was compared to determine if the treatment had any significant effect on the treatment group compared to the control group.

3. Observational study - sample population is measured. The researcher does not influence the population in any way or attempt to intervene in the study. There is no experimental manipulation. Instead, data is simply gathered and correlations are investigated.



Large Population

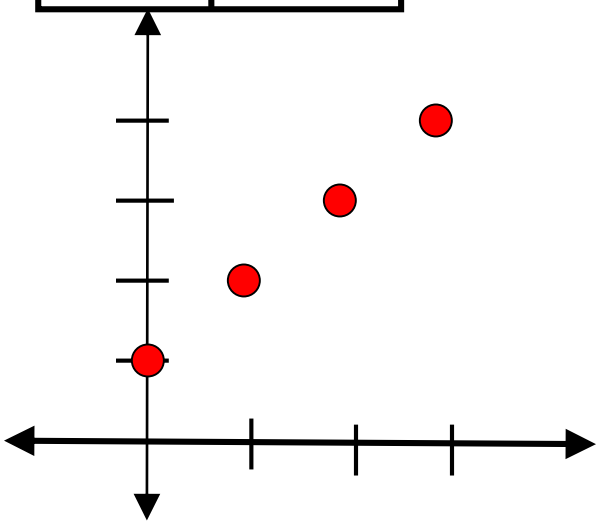


Compare by graphing

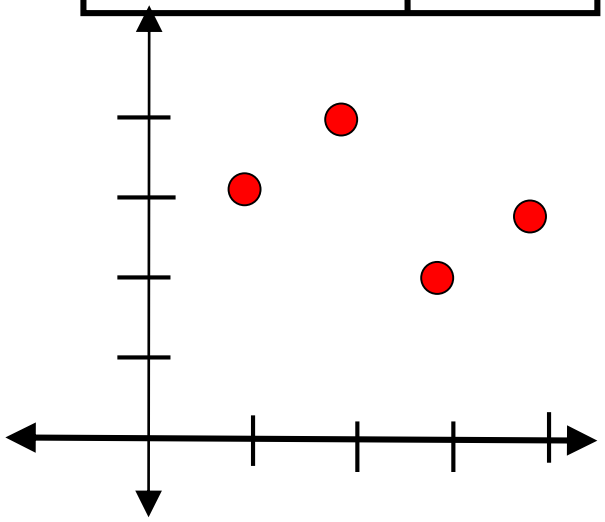
Correlation involves comparing trends in the data.

If one variable increases, does the other variable consistently increase or consistently decrease, or is there no observable trend in the second variable?

$\frac{\text{logs}}{\text{hr}}$	room temp
0	65
1	70
2	74
3	78



$\frac{\#\text{problems}}{\text{HW assignment}}$	Ave Test score
20	80
30	85
40	74
50	77



Comparing the Types of Statistical Studies

What is the difference between a sample study, an experimental study, or observational study?

Sample studies *estimate* a certain parameter of a population.

Experimental studies *compare* the measurements of pre- and post-treatment or *compare* the measurements of a group that has been treated to one that has not been treated.

Observational studies *compare* two different measurements (height vs. diet, etc.) in the population but no treatment is given.

Experimental studies and observational studies ***both*** *compare* two parameters of a population.

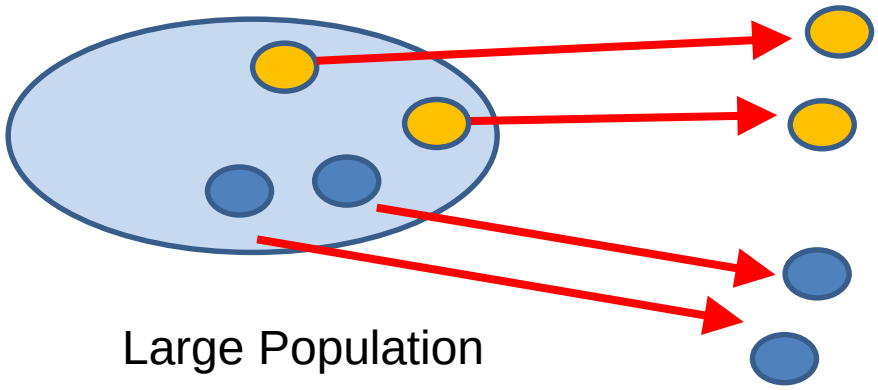
Sample studies one look at one parameter of a population.

What type of study is it?

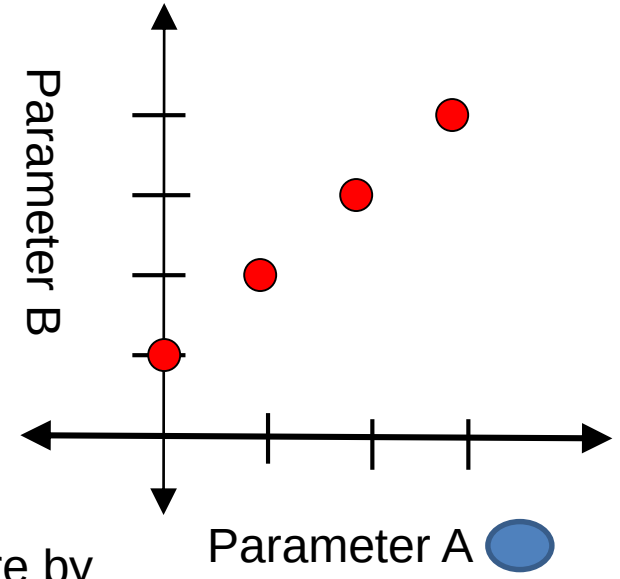
Imagine that you were interested in whether karaoke singers who are more animated would receive more applause from the audience. You visit a restaurant that has karaoke and observe karaoke singers during one evening. You rate each karaoke singer with respect to how animated the singer was. You rate the degree of applause that each singer received.

Observational study

3. Observational study - sample population is measured. The researcher does not influence the population in any way or attempt to intervene in the study. There is no experimental manipulation. Instead, data is simply gathered and correlations are investigated.



Large Population



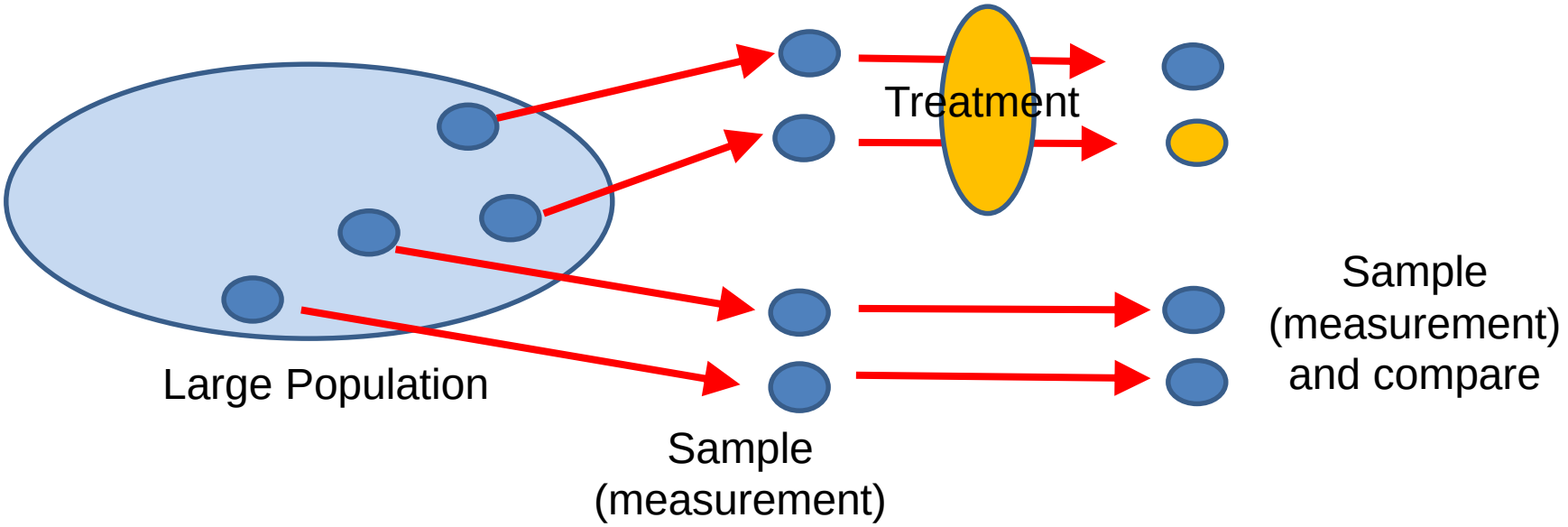
Compare by graphing

Which is it?
(Sample study, Observational study or Experimental?)

Giovanna has two possible driving routes to her new workplace, and she is having a hard time figuring out which is faster. The first route is shorter, but usually has more traffic than the second route.

Over a period of 5 weeks, she used the first route in the first half of the week (Monday, Tuesday, and Wednesday), and the second route in the second half of the week (Thursday and Friday). Each day she recorded how long it took her to get to work, and by the end of the 5 weeks she calculated the average driving duration for each route

“Controlled" experimental study:



The purpose of a sample study is to *estimate* a certain parameter of a population while observational studies and experiments *compare* two parameters of a population.

The study was designed to check the connection between the driving route and the driving duration. It's looking for a comparison, so this *isn't* a sample study. It could either be an experiment or an observational study.

In an experiment, one parameter of the population is changed to see the effect on the other parameter. In this case, the driving route was changed by Giovanna to each work day to assess its effect on the driving duration. Therefore, this is indeed an *experimental study.*