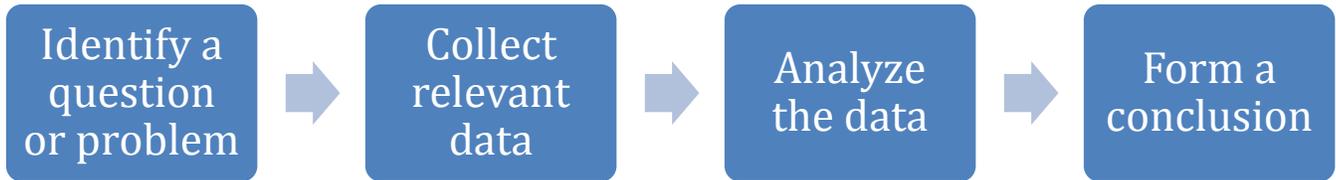


Chapter 5 Group Activity - SOLUTIONS

5A: Fundamentals of Statistics

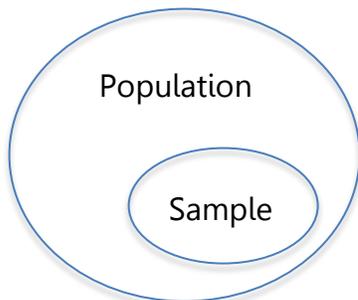
Group Activity

Statistics is the study of how to collect, analyze and make conclusions from data. Here is a diagram of the statistical process.

**Identify the Population, Sample, Parameter and Statistic**

1. For each scenario, a question or problem has been identified. Draw a diagram and describe the population, sample, parameter and statistic. Be specific with units for the parameter and statistic.

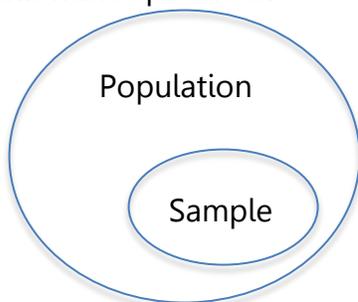
- a. A US insurance company wants to know whether households have two cars. A survey of 976 US households found that 32% of households have two cars.



Population – US households
Parameter – Percentage of US households that have two cars

Sample – 976 US households
Statistic – 32% of households in the survey have two cars.

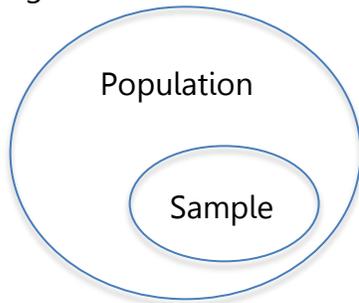
- b. A local toy store owner wants to know how much money kids in Portland get as an allowance per week.



Population – Kids in Portland
Parameter – Average dollar value that kids in Portland get for allowance per week

Sample – Some number of kids in Portland
Statistic – The average dollar value that kids in the sample get for allowance.

c. A beverage company wants to see if people in the United States liked their new logo.



Population – people in the US
Parameter – percentage who like the new logo

Sample – Some number of people in the US
Statistic – The percentage of people in the sample who like the new logo.

Sampling Methods

2. Identify the sampling method used:

a. All the Redland High School students were assembled in the gym, and then separated into groups by grade. Each student was assigned a number, and 25 numbers were randomly drawn from the students in each grade. They were asked if they planned to go to college.

Stratified

b. You are standing outside of the grocery store and stop every third person leaving to ask if they purchased milk.

Systematic

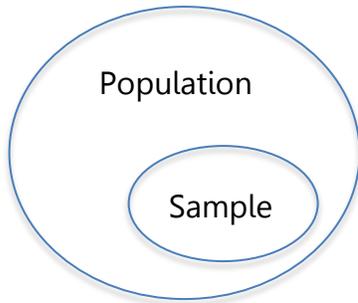
c. At his book club meeting, Dan asked each member if they had seen the movie made from their current book.

Convenience

d. Instead of the method in part a, the administrator at Redland High School chose 100 student numbers randomly and asked those students to report to the office for the survey.

Simple Random Sample

3. You are a statistical consultant and you have been hired to collect data on employee satisfaction. The company has 1000 employees at two buildings, East and West. There are 4 departments and 5 levels of employees. Identify the population, sample, parameter and statistic, then describe a sampling method.



Population – Employees in the company

Parameter – Percentage of all employees who are satisfied at work

Sample – A small group of employees

Statistic – The percentage of the sample who are satisfied at work

Stratified sampling makes the most sense here, by building, department, level or all three!

4. For each of the following, identify the **type of study** and fill in the corresponding boxes. Leave the remaining boxes blank.

Study	Observational	Observational: Case study		Experimental		
		Controls	Cases	Control Group	Treatment Group	Blinding
Over a 6-month period, among 100 people with bipolar disorder, patients given a high dose of omega-3 fats improved more than those given a placebo. Patients didn't know which they were given, but the experimenters knew.				People with bipolar disorder receiving the placebo	People with bipolar disorder receiving high dose of omega-3 fats	Single-blind (patients don't know but those giving the pills do)
A National Cancer Institute study of 716 melanoma patients and 1014 cancer-free patients found that those having a single large mole had twice the risk of melanoma.		Patients without a single large mole	Patients with a single large mole			
Over the period of one-year, researchers determined which airline had the lowest percentage of canceled flights.	Observational (not a case-study)					
1000 people were randomly separated into two groups—one group was assigned to exercise for 45 minutes daily, and the other group was instructed not to exercise. The participants in the exercise group reported falling asleep within 15 minutes of going to bed, but those not exercising laid awake for at least 30 minutes.				People not exercising	People exercising	No Blinding (participants know if they are in the exercise group)
500 patients with migraine headaches receive a shot to see if it helps reduce the frequency of the headaches. The nurses giving the shots write down the code on each syringe and the patient name. The syringes look identical, but half contain the new medicine and half contain saline.				Patients receiving saline (no medicine)	Patients receiving new medicine	Double-blind (neither patients nor nurses administering know which syringes have medicine)

Confidence Intervals

5. The table below lists results for a Gallup poll conducted via random telephone interviews in October 2015. Assume that the margin of error is 4 percentage points.

a. Find the confidence interval for each category.

	American adults	18-34 years old	35-49 years old	50-64 years old	65+ years old
Support Legalizing the use of Marijuana	58%	71%	64%	58%	35%
Confidence Interval	54% - 62%	67% - 75%	60% - 68%	54% - 62%	31% - 39%

b.
For

which age group(s) can you claim that over half of the people support legalization of marijuana? Explain.

For all age groups except those 65 and older, we can claim that over half of them support legalization of marijuana.

c. For which age group(s) can you claim that the majority of people do not support legalization of marijuana? Explain.

For those 65 and older, we can claim that the majority do not support legalization of marijuana.

d. Explain why the percentage for all American adults isn't exactly the average of all the categories.

The average (mean) for the four age categories is $(71\% + 64\% + 58\% + 35\%) / 4 = 57\%$ and the stated overall percentage for American adults is close at 58%, but not exact. This is because there is not the same number of people in each age category.

6. Go to Gallup.com or pewresearch.org and look up a poll about students. Give the percentage, margin of error and confidence interval. What do you think about the survey methods used?

One example is <https://news.gallup.com/poll/229085/college-students-say-campus-climate-deters-speech.aspx>

61% +/- 2% said in 2017 that their campus climate prevents them from saying what they believe, for a confidence interval of (59%, 63%). This is an increase from 2016, 54% +/- 2% or (52%, 56%).

5B: Types of Bias

7. In each situation, identify a potential source of bias. You will use each type from the class prep once.

- a. Trident gum did a survey of dentists to see how many would recommend their gum.

Self-interest study

- b. A boss asks their employees if they have taken drugs in the last week.

Lack of anonymity

- c. A researcher on a phone survey asks, "Do you plan to vote for the school district bond or would you rather see our schools crumble?"

Loaded question

- d. A survey was given to a random sample of students but three students didn't return the survey.

Non-response bias

- e. A survey about PCC student's experiences in Math 105 was given to students at Rock Creek and Sylvania.

Sampling bias (They did not include Cascade and SE)

- f. A survey asks people for their weight.

Response bias. Depending on the context, people may not give their true weight

- g. A broadcast email was sent to all PCC students with a satisfaction survey.

Voluntary response bias

5C: Statistical Tables and Graphs

Please complete the anonymous survey online. The link is in D2L. This will give our class some data to work with.

Types of Data

8. For each question in the survey, determine whether the data is qualitative or quantitative.

1. How do you identify your gender? **Qualitative**
2. What is your age? **Quantitative**
3. How many credits are you taking this term? **Quantitative**
4. What is your main mode of transportation to campus? **Qualitative**
5. How long is your pinky finger to the nearest half-centimeter? **Quantitative**

Frequency Tables and Graphs of Qualitative Data

9. For one of the sets of qualitative data, make a frequency table using a spreadsheet. Use your table to make a pie chart and a bar chart. Write a sentence or two about any patterns or observations you make from the graph.

Frequency Table for Gender

Male	
Female	
Non-binary	

Frequency Table for Transportation

Bus	
Walk	
Bike	
Drive Self	
Ride with another	
Other	

Quantitative Variables and Histograms

10. For the pinky length data, make a histogram by hand and/or using a spreadsheet. Write a sentence or two about any patterns or observations you make from the graph.

Time Series Plots

11. The data below show the number of people diagnosed with cancer each year. Make a time series plot in Excel. Write a sentence or two about any patterns or observations you make from the graph.

United States and Puerto Rico Cancer Statistics, 1999-2013 Incidence

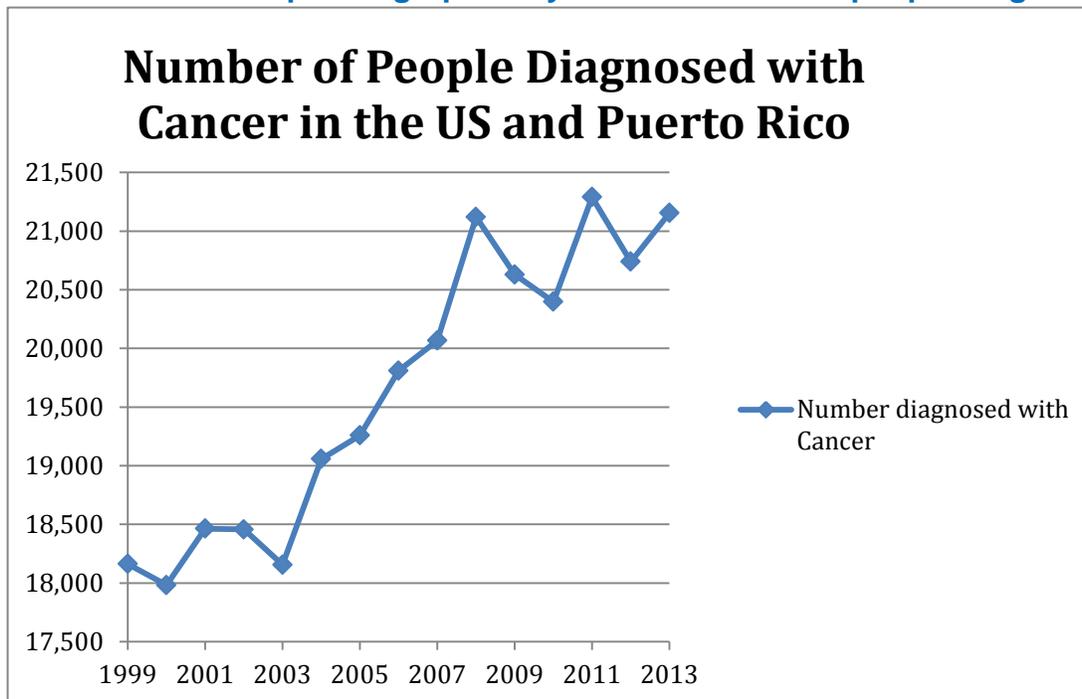
Year Number of people diagnosed with cancer

1999	18,162
2000	17,982
2001	18,463
2002	18,457
2003	18,155
2004	19,058
2005	19,258
2006	19,811
2007	20,067
2008	21,120
2009	20,630
2010	20,398
2011	21,290
2012	20,739
2013	21,155

Source: <https://wonder.cdc.gov/controller/datarequest/D121>

Show Cara your graphs when completed.

Note: Use a scatterplot to graph the year and number of people diagnosed.



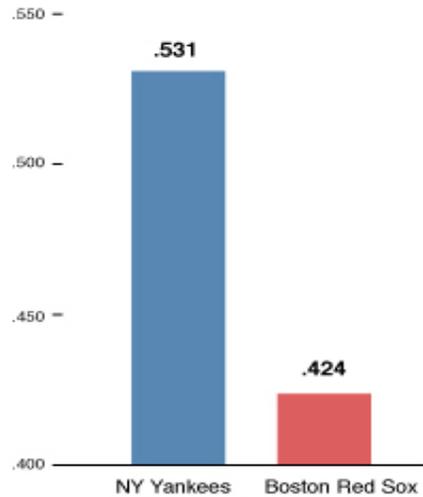
12. What is wrong with these graphs?

a. Ticket Prices



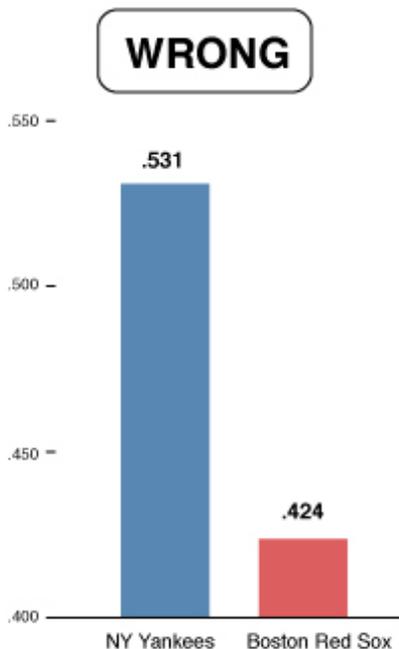
There is a visual distortion because instead of using bars, the balls increase in height and width. The baseball ticket is twice as much, but the area appears to be 4-8 times as much.

b. Percentage of Victories



Not starting at zero can make differences look larger. The difference between 42% and 53% is not that large. The scale should start at zero

Percentage of victories



Percentage of victories

