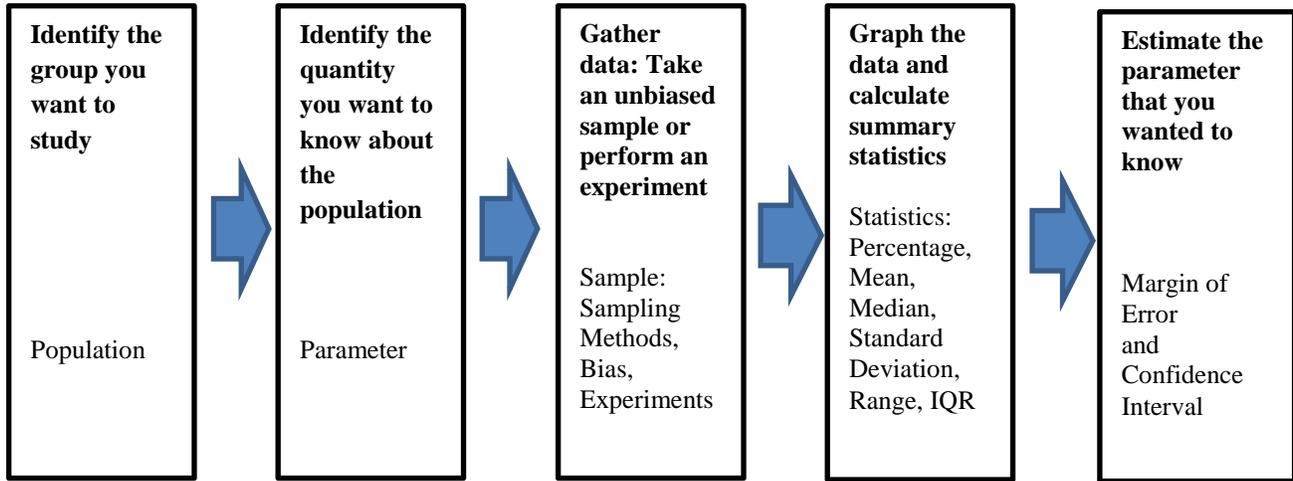


Chapter 3 Group Activity - SOLUTIONS

Group Activity

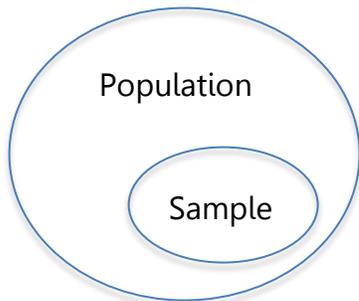
3.1 Overview of the Statistical Process



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 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.

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. In each situation, identify a potential **type 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

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

Study	Observational	Observational: Case study		Experiment		
		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.

3.2 Describing Data

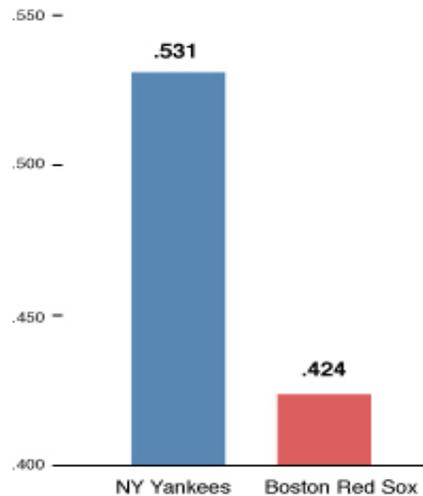
6. What is wrong with these graphs?

a. Ticket Prices



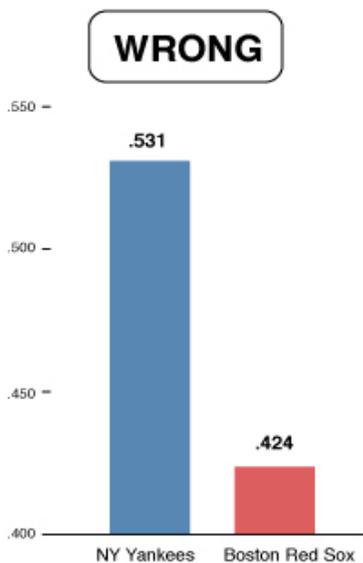
There is a visual distortion because instead of using bars, the balls increase in height and width. The baseball ticket is twice as much, but the area appears to be 4-8 times as much.

b. Percentage of Victories

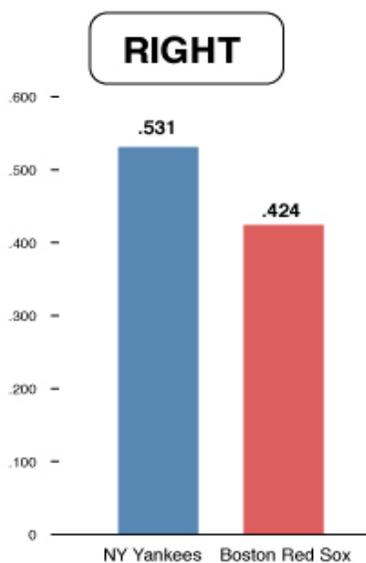


Not starting at zero can make differences look larger. The difference between 42% and 53% is not that large. The scale should start at zero.

Percentage of victories



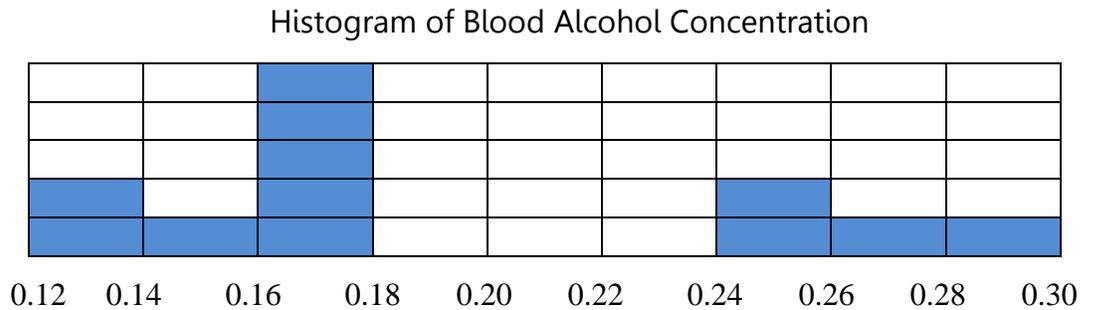
Percentage of victories



7. The following data are blood alcohol concentrations of 12 drivers involved in fatal crashes (data from the U.S. department of justice). We will analyze the shape, center and spread of this data, and whether there are any outliers.

a. On the first grid, make a histogram of the data using a bin-width of 0.02. Label your axes.

Blood alcohol concentration of drivers in fatal crashes
0.27
0.17
0.17
0.16
0.13
0.24
0.29
0.24
0.14
0.16
0.12
0.16



b. What is the shape of the histogram? If you are not sure yet, compare the mean and the median first.

The shape is bimodal and skewed to the right because the mean is greater than the median.

c. Find the mean, median and mode, include units.

Put in order: 0.12, 0.13, 0.14, 0.16, 0.16, 0.16, 0.17, 0.17, 0.24, 0.24, 0.27, 0.29

The mean is 0.1875 blood alcohol concentration

The median is 0.165 blood alcohol concentration

The mode is the center of the highest peaks at 0.17 blood alcohol concentration

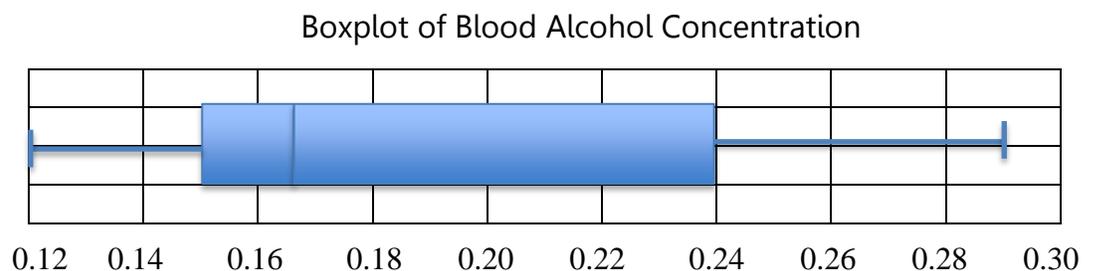
d. Find the 5-number summary, IQR and range, including units.

5-number summary: 0.12, 0.15, 0.165, 0.24, 0.29 blood alcohol concentration

IQR = $Q3 - Q1 = 0.24 - 0.15 = 0.09$ blood alcohol concentration

Range = $0.29 - 0.12 = 0.17$ blood alcohol concentration

e. Use the 5-number summary to draw a boxplot on the second grid below. Notice the horizontal scales match on both graph so we can see how they compare.



f. Do you think there are any outliers? Why or why not? **There do not seem to be any outliers but there is an unusual gap.**

Calculating Standard Deviation, s

d. Using your mean rounded to two decimal place, find the standard deviation, including units. The variable n refers to the number of data values.

Mean = 0.19 blood alcohol concentration, n = 12 drivers

Test Score (points)	Deviation from the mean	Squared deviation
0.27 - 0.19	0.08	0.0064
0.17 - 0.19	- 0.02	0.0004
0.17 - 0.19	- 0.02	0.0004
0.16 - 0.19	- 0.03	0.0009
0.13 - 0.19	- 0.06	0.0036
0.24 - 0.19	0.05	0.0025
0.29 - 0.19	0.10	0.0100
0.24 - 0.19	0.05	0.0025
0.14 - 0.19	- 0.05	0.0025
0.16 - 0.19	- 0.03	0.0009
0.12 - 0.19	- 0.07	0.0049
0.16 - 0.19	- 0.03	0.0009
Sum of the squared deviations (numerator)		0.0359

$$s = \sqrt{\frac{\sum (x - \text{mean})^2}{n - 1}}$$

$$= \sqrt{\frac{0.0359}{12 - 1}}$$

≈ 0.0571 blood alcohol concentration