Chapter 1 Review Sheet with Answer Key

Vocabulary Terms from sections 1A – 1D

Data: Information in all forms.

Categorical data: Also called qualitative data. Data in the form of labels that tell us something about the people or objects in the data set. For example, the country they live in, occupation, or type of pet.

Quantitative data: Data in the form of numbers that measure or count something. They usually have units and taking an average makes sense. For example, height, weight, salary, or the number of pets a person has.

Population: The collection of all people or objects to be studied.

Census: Collecting data from everyone in a population.

Sample: Collecting data from a small subgroup of the population.

Random: When everyone in the population has a chance to be included in the sample.

Simple Random Sample: Sample data in which individuals are selected randomly. This method tends to minimize sampling bias and is generally considered a good way to collect data.

Convenience Sample: Sample data that is collected in a way that is easy or convenient. This method tends to have a significant amount of sampling bias and is generally considered a bad way to collect data.

Voluntary Response Sample: Sample data that is collected by putting a survey out into the world and allowing anyone to fill it out. This method tends to have a significant amount of sampling bias and is generally considered a bad way to collect data.

Cluster Sample: Sample data that collects data from groups of people in a population instead of one at a time. The groups should be chosen randomly to avoid sampling bias.

Stratified Sample: Sample data used to compare two or more groups or compare two or more populations. The individuals from each group should be chosen randomly to avoid sampling bias. For example, we may take a random sample of people living in Palmdale, CA and another random sample of people living in Valencia, CA and use the data to compare the average salaries.

Systematic Sample: Sample data that is collected with some type of system like choosing every twentieth person on a list.

Bias: When data does not represent the population.

Sampling Bias: A type of bias that results from collecting data without using a census or random sample. The method of collecting is flawed. For example, using convenience or voluntary response method to collect the data. We can minimize this bias by collecting the data with a census or random sample.

Question Bias: A type of bias that results when someone phrases the question or gives extra information with the goal of tricking the person into answering a certain way. We can minimize this bias by phrasing our questions in a neutral way and not attempt to sway the person giving data.

Response Bias: A type of bias that results when people giving the data do not answer truthfully or accurately. To minimize this bias, we should collect the data anonymously and assure the person giving the data that the data will be used for scientific purposes and will not be released.

Non-response Bias: A type of bias that results when people refuse to participate or give data. To minimize non-response bias, you may give an incentive like a gift card to encourage people to give data.

Deliberate Bias: A type of bias that results when the people collecting the data falsify the reports, delete data, or decide to not collect data from certain groups in the population. To minimize deliberate bias, the people collecting and analyzing the data need to have good ethics. They should not falsify reports, delete data or leave out groups from the population.

Experimental Design: A scientific method for controlling confounding variables and proving cause and effect.

Observational Study: Collecting data without controlling confounding variables. This type of data cannot prove cause and effect.

Explanatory Variable: The independent or treatment variable. In a cause and effect experiment, this is the cause variable.

Response Variable: The dependent variable. In a cause and effect experiment, this the variable that measures the effect.

Treatment Group: The group of people or objects that has the explanatory variable. In an experiment involving medicine, this would be the group that receives the medicine.

Control Group: The group of people or objects that is used to compare and does not have the explanatory variable. In an experiment involving medicine, this would be the group that receives the placebo.

Confounding Variables: Also called lurking variables. Other variables that might influence the response variable other than the explanatory variable being studied.

Random assignment: A process for creating similar groups where you take a group of people or objects and randomly split them into two or more groups.

Placebo Effect: The capacity of the human brain to manifest physical responses based on the person believing something is true.

Placebo: A fake medicine or fake treatment used to control the placebo effect.

Vocabulary Terms from sections 1E – 1G

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Percentage: A statistic calculated from categorical data that measures the part out of 100.

Proportion: The decimal equivalent to a percentage.

Sample Size (n): Also called the total frequency. This is the total number of people or objects represented in the sample data. If we collected data from 35 people, then the sample size would be n = 35.

Mean: A measure of center or average for quantitative data that balances the distances. The mean average is only accurate if the quantitative data is normal (bell shaped). Hence, the mean average is the center or average used for normal quantitative data.

Median: A measure of center or average for quantitative data that is found by finding the center of the data when the data values are in order. The median is the most accurate center or average when the data is skewed left, skewed right, or not normal.

Mode: The number or numbers that appear most often in a quantitate data set. The mode is used as a measure of center or average.

Midrange: A quick measure of center or average that is half way between the min and max of a quantitative data set. It is generally not very accurate, but easy to calculate.

Standard Deviation: The most accurate measure of typical spread for normal (bell shaped) quantitative data. The standard deviation measures how far typical values are from the mean on average. The standard deviation is only accurate if the quantitative data is normal (bell shaped).

Variance: A measure of spread used in ANOVA testing. The variance is the square of the standard deviation and is only accurate when data is normal (bell shaped).

Interquartile Range (IQR): The most accurate measure of spread for skewed or non-normal data. The IQR measures how far typical values are from each other in skewed or non-normal data. IQR is calculated by subtracting the Third Quartile (Q3) minus the First Quartile (Q1).

Range: A quick measure of spread that measures the distance between the max and min of a quantitative data set. It is easy to calculate (Max – Min), but is not an accurate measure of typical spread, since it does not involve typical values in the data.

First Quartile (Q1): The divider that approximately 25% of the quantitative data values are less than. Q1 is the bottom range of typical values for skewed or non-normal data. Typical values are between Q1 and Q3 in skewed or non-normal data. Q1 is considered a measure of position.

Third Quartile (Q3): The divider that approximately 75% of the quantitative data values are less than. Q3 is the top range of typical values for skewed or non-normal data. Typical values are between Q1 and Q3 in skewed or non-normal data. Q3 is considered a measure of position.

Max: The largest number in a quantitative data set. Considered a measure of position.

Min: The smallest number in a quantitative data set. Considered a measure of position.

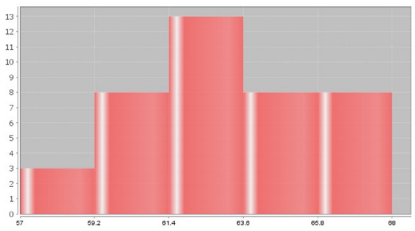
Categorical Data Analysis (Section 1E)

- **To convert a decimal proportion into a percentage => Multiply by 100 and put on the % symbol.** (This will move the decimal two places to the right.)
- **To convert a percentage into a decimal proportion => Remove the % symbol and Divide by 100.** (This will move the decimal two places to the left.)
- **To calculate the proportion for each categorical variable: Proportion =** $\frac{x}{n} = \frac{Amount\ (\#\ of\ successes)}{Total\ Frequency\ (Sample\ Size)}$ (StatKey calculate counts and proportions for you.)
- **Round Proportions to the thousandths place.** (Three numbers to the right of decimal point. StatKey round to the thousandths place.)

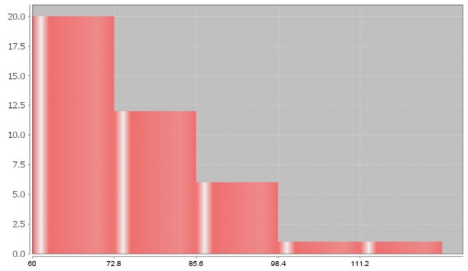
Quantitative Data Analysis (Sections 1F & 1G)

Shapes

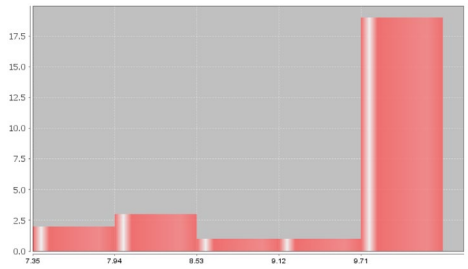
Normal (Bell Shaped, Unimodal and Symmetric)



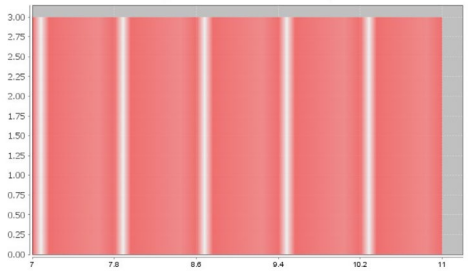
Skewed Right (Positively Skewed)



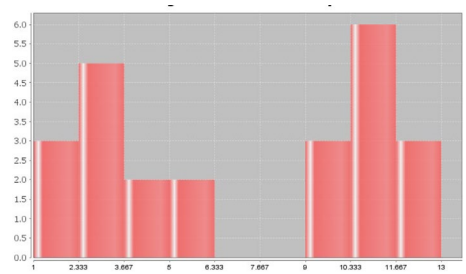
Skewed Left (Negatively Skewed)



Uniform



Bimodal



Shape determine what statistics are accurate!

Normal Quantitative Data

Center (Average): Mean

Typical Spread: Standard Deviation

Typical Values Between: Mean – Standard Deviation and Mean + Standard Deviation
(One Standard Deviation above and below the mean, Middle 68% of the data)

High Outliers: Data Values ≥ Mean + (2 x Standard Deviation)
(Two standard deviations above the mean, Top 2.5% of the data)

Low Outliers: Data Values ≤ Mean − (2 x Standard Deviation)
(Two standard deviations below the mean, Bottom 2.5% of the data)

Skewed or Non-normal Quantitative Data

Center (Average): Median

Typical Spread: Interquartile Range (IQR)

Typical Values Between: 1st quartile (Q1) and 3rd quartile (Q3)
(Middle 50% of data values)

High Outliers: Data Values ≥ Q3 + (1.5 x IQR) *Automatically calculated in Box-Plot!*

Low Outliers: Data Values ≤ Q1 − (1.5 x IQR) *Automatically calculated in Box-Plot!*

Chapter 1 Review Problems

Problems from Sections 1A – 1D

1. Tell if the following data is categorical or quantitative and explain why.

    a) The types of cars in the different parking lots.

    b) The average number of hours spent practicing ping-pong.

    c) Areas in North Dakota that have wild mustangs.

    d) Each person is asked if he or she wear glasses, contacts, neither, or both.

    e) The average speed of racecars at the Indianapolis 500.

    f) Exam scores for various students on a history exam.

2. Jim wants to know how much money the average working COC student makes. Describe how Jim could use each of the following techniques to collect data. For each technique, will there be a significant amount of sampling bias or not too much sampling bias?

a) Systematic

b) Voluntary Response

c) Random Sample

d) Convenience Sample

e) Cluster Sample

f) Stratified Sample

g) Simple Random Sample

h) Census

3. Define the following key terms and give an example of each.

a) Population

b) Census

c) Sample

d) Random

e) Bias

f) Statistic

4. Describe and give an example of each of the following types of bias. Also state how a person collecting and analyzing data, can avoid these biases.

a) Sampling Bias

b) Question Bias

c) Response Bias

d) Deliberate Bias

e) Non-Response Bias

5. Rachael needs to do an experiment that will show that wearing nicotine patches cause a person to stop smoking. Set up the experiment for Rachael. What is the explanatory variable? What is the response variable? Write a description of the experiment and include the following. What are some confounding variables that she will need to control? How can Rachael control the confounding variables? Include a description of how Rachael use a double blind placebo to control the placebo effect. Describe the treatment group and the control group in the experiment.

6. Compare and contrast the similarities and differences between an experiment and an observational study. How can we tell if we should use an experiment or an observational study?

Problems from Sections 1E – 1G

7. Explain the following.

    a) Explain how to round a decimal to a given place value.
    b) Explain how to convert a decimal proportion into a percentage.
    c) Explain how to convert a percentage into a decimal proportion.
    d) Explain how to calculate a percentage by using an amount and a total from categorical data.
    e) Explain how to calculate an estimated amount by using a percentage and a total from categorical data.

8. Convert the following proportions into percentages. Do not round your answer.

    a) 0.0722
    b) 0.0041
    c) 0.563
    d) 0.0005

9. Convert the following percentages into decimal proportions. Do not round your answer.

    a) 35.9%
    b) 4.823%
    c) 0.026%
    d) 0.389%

10. A company has 74 employees. Of those employees 11 are managers, 27 are full-time employees and 36 are part-time employees. Use this information to answer the following questions.

    a) What proportion of the employees are managers? *(Round your answer to the thousandths place.)*
    b) What percentage of the employees are managers? *(Round your answer to the tenths place.)*
    c) What proportion of the employees are full-time employees? *(Round your answer to the thousandths place.)*
    d) What percentage of the employees are full-time employees? *(Round your answer to the tenths place.)*
    e) What proportion of the employees are part-time employees? *(Round your answer to the thousandths place.)*
    f) What percentage of the employees are part-time employees? *(Round your answer to the tenths place.)*
    g) Calculate the percent of increase between managers and full-time employees. Is there a significant difference between the percentages? Explain why.
    h) Calculate the percent of increase between full-time and part-time employees. Is there a significant difference between the percentages? Explain why.

11. According to an online article, approximately 60% of the voting population in the U.S. votes during a presidential election year. According to a census, there are approximately 41,743 people living in Saugus, CA. If 60% of them vote in the next presidential election, how many people do we expect to vote in Saugus?

12. Describe and draw a histogram for each of the following shapes.

    a) Normal
    b) Skewed Right
    c) Skewed Left
    d) Uniform
    e) Bimodal

13. Classify each of the following quantitative statistics as a measure of center, spread or position. Also, describe when that statistic should be used.

a) Q1
b) Mean
c) Variance
d) Standard Deviation
e) Minimum
f) Q3
g) Mode
h) IQR
i) Median
j) Range
k) Maximum
l) Midrange

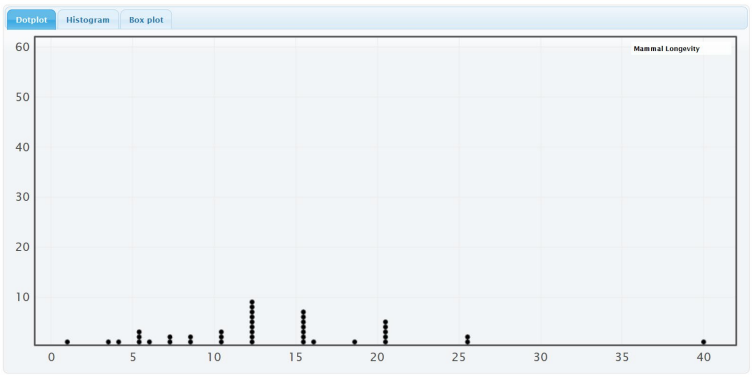
14. Answer each of the following questions about quantitative data analysis.

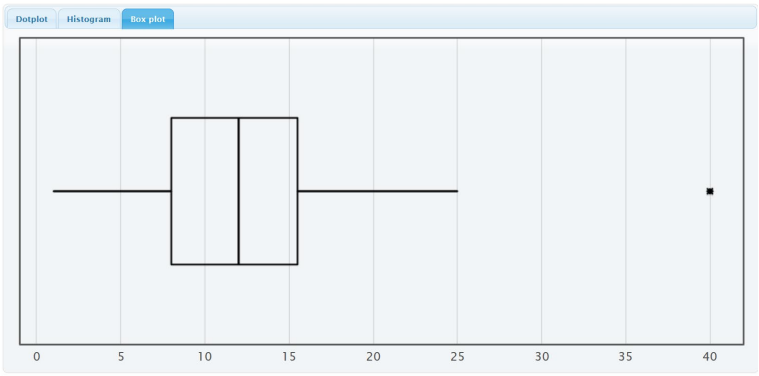
a) What measure of center (average) should we use if the data is normal?
b) What measure of center (average) should we use if the data is not normal?
c) What measure of spread (variability) should we use if the data is normal?
d) What measure of spread (variability) should we use if the data is not normal?
e) How do we find two numbers that typical values fall in between if the data is normal?
f) How do we find two numbers that typical values fall in between if the data is not normal?
g) What is the formula for the high outlier cutoff if the data is normal?
h) What is the formula for the high outlier cutoff if the data is not normal?
i) What is the formula for the low outlier cutoff if the data is normal?
j) What is the formula for the low outlier cutoff if the data is not normal?
k) How do we determine if a data value is an outlier when the data is normal?
l) How do we determine if a data value is an outlier when the data is not normal?

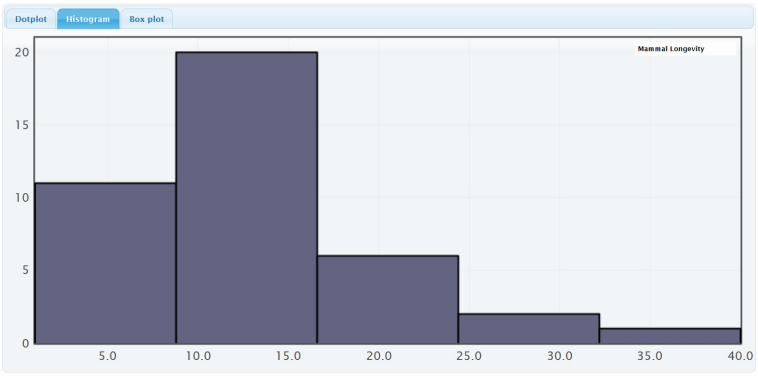
15. The following graphs and statistics were calculated with StatKey and describe the number of years mammals live. Use the graphs and statistics to answer the following questions.

Summary Statistics

Statistic	Value
Sample Size	40
Mean	13.150
Standard Deviation	7.245
Minimum	1
Q_1	8.000
Median	12.000
Q_3	15.500
Maximum	40







a) What is the data measuring and what are the units?
b) How many mammals are in the data set?
c) What is the shape of the data set?
d) What is the minimum value?
e) What is the maximum value?
f) What is the average (center)? *(Give the number and the name of the statistic used.)*
g) How much typical spread does the data set have?
(Give the number and the name of the statistic used.)
h) Find two numbers that typical values fall in between.
i) List all high outliers in this data set. If there are no high outliers, put "none".
j) List all low outliers in this data set. If there are no high outliers, put "none".

16. The following graphs and statistics were calculated with Statcato and describe the number of years employees have been employed at a company. Use the graphs and statistics to answer the following questions.

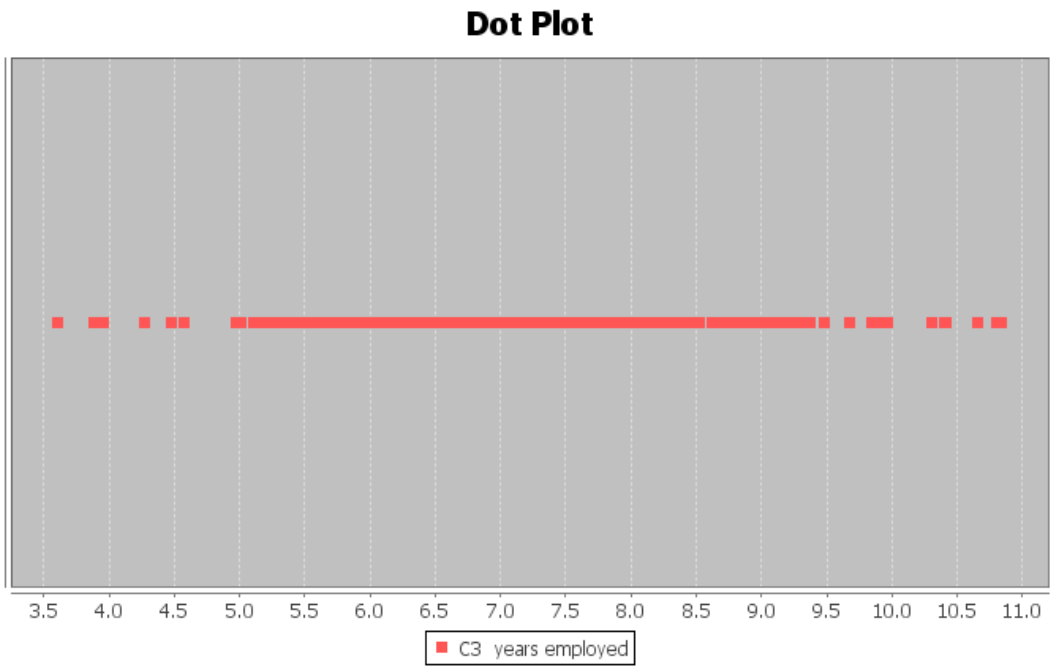
Descriptive Statistics

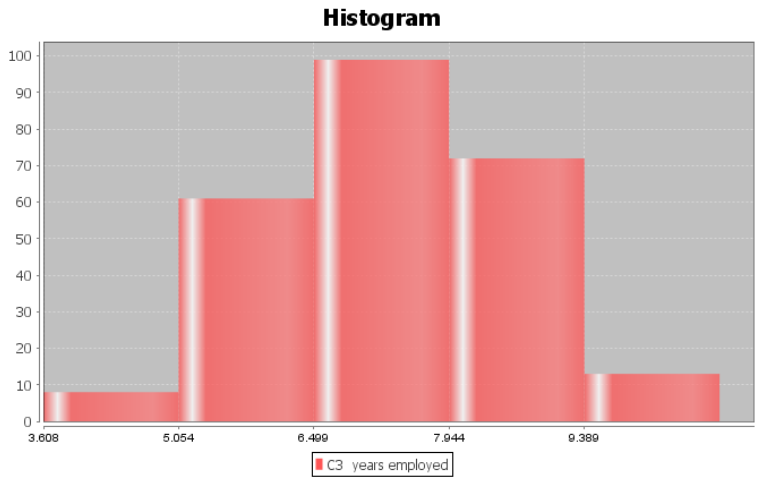
Variable	Mean	Standard Deviation
years employed	7.345	1.376

Variable	Q1	Median	Q3	IQR
years employed	6.4	7.35	8.3	1.9

Variable	Min	Max	Range
years employed	3.6	10.8	7.2

Variable	N total
years employ	253

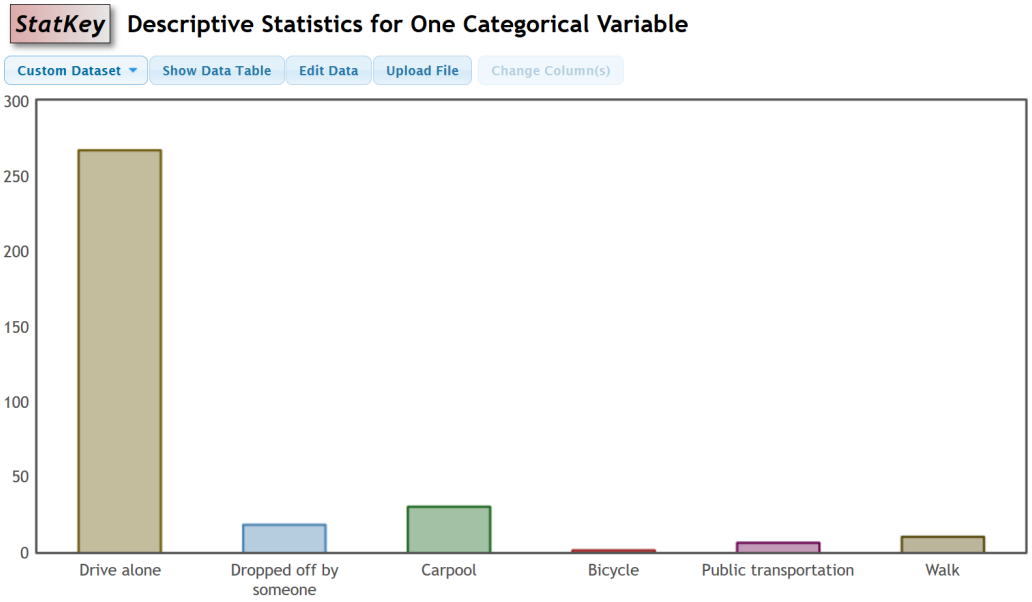




a) What is the data measuring and what are the units?
b) How many employees are in the data set?
c) What is the shape of the data set?
d) What is the minimum value?
e) What is the maximum value?
f) What is the average (center)? *(Give the number and the name of the statistic used.)*
g) How much typical spread does the data set have?
(Give the number and the name of the statistic used.)
h) Find two numbers that typical values fall in between.
i) Calculate the high-outlier cutoff. Give approximate values of the high outliers in this data set. If there are no high outliers, put "none".
i) Calculate the low-outlier cutoff. Give approximate values of the low outliers in this data set. If there are no low outliers, put "none".

17.

Statistics students were asked what mode of transportation they take to get to school. Use the following bar chart and statistics to answer the following.



Summary Statistics

	Count	Proportion
Drive alone	267	0.804
Dropped off by someone	18	0.054
Carpool	30	0.09
Bicycle	1	0.003
Public transportation	6	0.018
Walk	10	0.03
Total	332	1.000

a) What was the most popular mode of transportation?

b) What was the least popular mode of transportation?

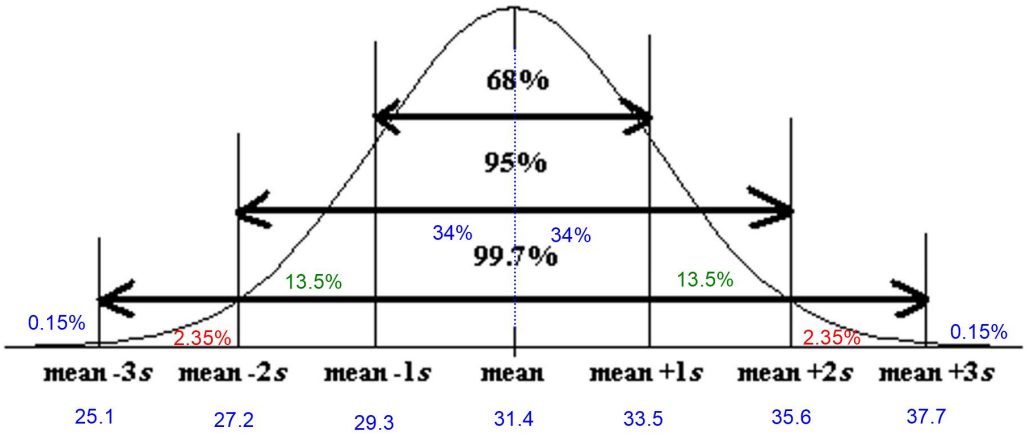
c) How many statistics students walked to school?

d) What proportion of statistics students were dropped off by someone? Do not calculate the answer. Use the table provided. Do not round the answer.

e) What percentage of the statistics students use public transportation? Do not calculate the answer. Use the table provided and convert the answer into a percentage. Do not round the answer.

18.

The salaries of employees at a company are normally distributed with a mean of 31.4 thousand dollars and a standard deviation of 2.1 thousand dollars. Use the Empirical Rule graph below to answer the following questions about this normal quantitative data.



a) What percent of the salaries are between 29.3 thousand dollars and 31.4 thousand dollars?

b) What percent of the salaries are 33.5 thousand dollars or more?

c) Typical salaries are between what two values?

d) What is the high outlier cutoff?

e) What is the low outlier cutoff?

f) What is the average salary for employees of this company?

Chapter 1 Review Sheet Answers

1.

a) Categorical since the data would consist of words.
b) Quantitative since it is numerical measurement data.
c) Categorical since the data would consist of words.
d) Categorical since the data would consist of words.
e) Quantitative since it is numerical measurement data.
f) Quantitative since it is numerical measurement data.

2.

a) Jim can ask every 5th student that walks into the COC cafeteria about their salary. This would have a significant amount of sampling bias.

b) Jim can put a survey on Facebook asking how money COC students make. This would have a significant amount of sampling bias.

c) Jim can have a computer randomly select student ID numbers and then track down those students whose ID numbers were selected and ask them their salary. This would have no sampling bias.

d) Jim can ask other students in his COC classes about their salary. This would have a significant amount of sampling bias since it is not a random sample.

e) Jim can randomly select 10 section numbers at COC, and then go to those classes and get data from everyone in the class. Since he chose the groups randomly, this would not have much sampling bias.

f) Jim could walk around the COC campus asking female students about their salary. Later he could walk around asking male students about their salary. Later he could compare the female and male student salaries. Since this method was not randomly selected, there would be a lot of sampling bias.

3.

Population: The collection of all people or objects to be studied. For example, a marine biologist could study all dolphins in the world.

Census: Collecting data from everyone in a population. This is the best way to collect data and minimizes sampling bias. For example, suppose our population of interest was the students at Valencia high school. We could collect data from every student at Valencia high school.

Sample: Collecting data from a small subgroup of the population. For example, if our population was all people in Palmdale, CA, we might collect data from fifty people in Palmdale.

Random: When everyone in the population has a chance to be included in the sample. Suppose our population is all COC students. We could have a computer randomly select student ID numbers and then collect data from those students.

Bias: When data does not represent the population. Asking your friends and family will not represent the population of all people in the world.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. Sample mean averages, sample standard deviations, or sample percentages would all be examples of statistics.

4.

Sampling Bias: A type of bias that results from collecting sample data that is not random or representative of the population. For example, if our population was all adults in California, and our sample consists of asking our friends and family. To limit this bias, we could take a random sample instead.

Question Bias: A type of bias that results when someone phrases the question or gives extra information with the goal of swaying the person to answer a certain way. Instead of asking a person's opinion about raising taxes, the person first gives a speech about how they think raising taxes is terrible. To limit this bias we could simply ask if the person is for raising or lowering taxes and not give any extra information.

Response Bias: A type of bias that results when people do not answer truthfully or accurately. Asking people how much they weigh in pounds will result in many people lying about the answer. Instead of asking people, we could weigh them on a scale and assure them the data will not be released.

Deliberate Bias: A type of bias that results when the people collecting the data falsify the reports, delete data, or decide to not collect data from certain groups in the population. A common deliberate bias is to delete all of the data that makes your company look bad. We could avoid this bias by not deleting data or falsifying reports. Use the data to improve the company.

Non-response Bias: A type of bias that results when people refuse to participate or give data. When calling random phone numbers to collect data, many people will refuse to answer. To limit this bias, we may leave a message asking them to call us back and offering a gift card if they do.

5. Rachael will need a group of volunteers who want to participate in the experiment. She will need to randomly assign the volunteers into two groups. One group will be the treatment group and receive actual nicotine patches. The other group will be the control group and receive a fake patch (placebo). The placebo patch and the real patch should look identical. Patches should be given to patients using a double blind approach. No volunteer in the experiment will know if they are getting the real patch or a placebo. Also those directly giving the patch will not know either. This will control the placebo effect. Randomly assigning the groups will make them alike in many confounding variables. Rachael may also exercise direct control and manipulate the groups so that they are even more alike. There are many confounding variables including the level of addiction, the number of cigarettes smoked previously, genetics, age, gender, stress, job, etc. Answers may vary. Random assignment should control these confounding variables. If the experiment shows that those with the patch have a significantly higher percentage of quitting smoking, then it will prove that using the patch causes a person to quit smoking.

6.

An experiment creates two or more similar groups with either random assignment or using the same people twice. The similar groups control confounding variables and prove cause and effect. An observational study does not create similar groups and does not control confounding variables. An observational study just collects data and analyzes it, so it cannot prove cause and effect.

Experiment Example: Suppose we want to prove that drinking alcohol causes car accidents. We can have a group of volunteers that wish to participate. We create a driving course with cones. All of the volunteers drive the course sober and we keep track of the number of cones struck. All volunteers drive the same car, with no other distractions (no phones or radio). Then we allow the volunteers to drink alcohol until they all have similar blood alcohol content. Then they can re-drive the course and we keep track of the number of cones struck. If the number of cones is significantly more in the drunk drivers, we have proven that drinking alcohol causes car accidents.

Observational Study Example: Suppose we collect data on car accidents and how many of them involved drunk driving. There are many things that influence having a car accidents other than alcohol, so this data would not prove cause and effect.

7.

a) Identify the place value you wish to round. Look at the number to the right of the place value. If the number is 5 or above, add 1 to the place value and cut off the rest of the decimal. If the number is 4 or less, leave the place value alone and cut off the rest of the decimal.

b) To convert a decimal proportion into a percentage, simply multiply the decimal by 100 and add on the "%" sign.

c) To convert a percentage into a decimal proportion, remove the "%" sign, and divide the percentage by 100.

d) To calculate a percentage divide the amount by the total.

e) To estimate an amount, convert the percentage into a decimal proportion and multiply the proportion by the total. Round the answer to the ones place.

8.

a) 7.22%

b) 0.41%

c) 56.3%

d) 0.05%

9.

a) 0.359

b) 0.04823

c) 0.00026

d) 0.00389

10.

a) 11/74 ≈ 0.149

b) 0.149 x 100% = 14.9%

Approximately 14.9% of the company are managers.

c) 27/74 ≈ 0.365

d) 0.365 x 100% = 36.5%

Approximately 36.5% of the company are full-time employees.

e) 36/74 ≈ 0.486

f) 0.486 x 100% = 48.6%

Approximately 48.6% of the company are part-time employees.

g) Percent of Increase = (0.365 – 0.149) / 0.149 ≈ 145.0% increase. This seems to be a significantly large percent of increase so the percentage of full-time employees seems significantly higher than the percentage of managers. The difference also seems to be practically significant since there are 16 more full time employees than managers and the whole company is 74 total.

h) Percent of Increase = (0.486 – 0.365) / 0.365 ≈ 33.2% increase. This seems to be a large percent of increase so the percentage of part-time employees seems significantly higher than the percentage of full-time employees. The difference may not be practically significant since there are only 9 more part-time employees than full-time.

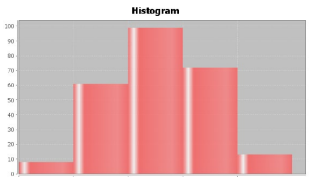
11.

60% = 0.6

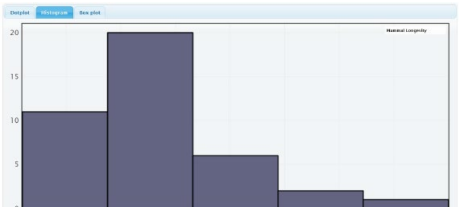
Estimated Amount = 0.6 x 41743 ≈ 25,046 voters in Saugus.

12.

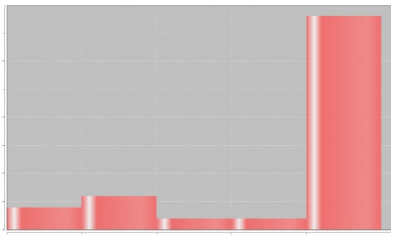
a) A normal or normally distributed histogram is unimodal and symmetric. This means that we expect the highest bar or bars to be in the middle with smaller and smaller bars as we go away from the middle. The left and right tails will be approximately the same length.



b) A skewed right or positively skewed histogram will have the highest bar or bars on the far left of the graph. It will have very few bars to the left of the center and many bars to the right of the center. Therefore the right tail will look much longer than the left tail.



c) A skewed left or negatively skewed histogram will have the highest bar or bars on the far right of the graph. It will have very few bars to the right of the center and many bars to the left of the center. Therefore, the left tail will look much longer than the right tail.



13.

a) The first quartile (Q1) is a measure of position. It is used to analyze typical values when data is skewed or not normal.

b) The mean is a measure of center. It is the primary center or average when the data is normal.

c) The variance is a measure of spread. It is used when the data is normal.

d) The standard deviation is a measure of spread. It is the primary measure of spread when the data is normal.

e) The minimum value is a measure of position. It can sometimes be an outlier and is used in all quantitative data regardless of shape.

f) The third quartile (Q3) is a measure of position. It is used to analyze typical values when data is skewed or not normal.

g) The mode is a measure of center. It is often used in business applications or any time we wish to know the value or values that appear most often.

h) The interquartile range (IQR) is a measure of spread. It is the primary measure of spread when the data is skewed or not normal.

i) The median or 50^{th} percentile or 2^{nd} quartile (Q2) is a measure of center. It is the primary measure of center or average when the data is skewed or not normal.

j) The range is a measure of spread. It is usually used when someone wants a quick easy to calculate measure of spread. It does not represent typical spread.

k) The maximum value is a measure of position. It can sometimes be an outlier and is used in all quantitative data regardless of shape.

l) The midrange is a measure of center. It is usually used when someone wants a quick easy to calculate center or average. It may not be a very accurate average since it is often based on outliers.

14.

a) We should use the mean average when the data is normal.

b) We should use the median average when the data is not normal.

c) We should use the standard deviation as our main measure of typical spread when data is normal.

d) We should use the interquartile range (IQR) as our main measure of typical spread when data is not normal.

e) When the data is normal, add and subtract the mean and standard deviation. Typical values will fall between $\bar{x}$ − s and $\bar{x}$ + s .

f) When data is not normal, typical values will fall between Q_1 and Q_3.

g) To calculate the unusually high cutoff for normal data, multiply the standard deviation by two and then add it to the mean ($\bar{x}$ + 2s).

h) To calculate the unusually high cutoff for non-normal data, multiply the IQR by 1.5 and add it to Q_3. (Q_3 + (1.5xIQR)).

i) To calculate the unusually low cutoff for normal data, multiply the standard deviation by two and then subtract from the mean ($\bar{x}$ − 2s).

j) To calculate the unusually low cutoff for non-normal data, multiply the IQR by 1.5 and subtract it from Q_1. (Q_1 − (1.5xIQR)).

k) To find low outliers in a normal data set, calculate the unusual low cutoff $\bar{x}$ − 2s and look for any data values that are lower than the low cutoff. To find high outliers in a normal data, calculate the unusual high cutoff $\bar{x}$ + 2s and look for any data values that are higher than the high cutoff.

l) To find low and high outliers in a skewed or non-normal data set, create a boxplot and look for any stars, circles or triangles outside of the whiskers.

15.

a) The data is measuring the age of mammals. The units are in years.

b) Sample size = 40. There are 40 mammals in the data set.

c) The data is skewed right.

d) The youngest mammal was 1 year old.

e) The oldest mammal was 40 years old.

f) Since the data was not normal, we will use the median average. The average age of the mammals is 12 years.

g) Since the data was not normal, we will use the interquartile range to measure the typical spread. IQR = Q3 – Q1 = 15.5 – 8 = 7.5 years. So typical mammal ages in this data set are within 7.5 years of each other.

h) Since the data was not normal, typical values will fall between Q1 and Q3. So typical mammal ages in this data set are between 8 years and 15.5 years.

i) The box plot shows that there is one high outlier at 40 years.

j) The box plot shows that there is no low outliers.

16.

a) The data is measuring the amount of time employees have been with the company. The units are years.

b) N total = 253. There are 253 employees in the data set.

c) The data appears normal.

d) The employee that has been with the company the shortest amount of time is 3.6 years.

e) The employee that has been with the company the longest amount of time is 10.8 years.

f) Since the data is normal, we will use the mean average. The average time that employees have been with this company is 7.345 years.

g) Since the data is normal, we will use the standard deviation as our measure of typical spread. So typical employee times with the company are within 1.376 years of the mean.

h) Since the data is normal, the high outlier cutoff is the mean + (2 x standard deviation) = 7.345 + (2 x 1.376) = 10.097. So any employees that have been with the company 10.097 years or moreis considered an unusually large amount of time. The dot plot shows five employees that have been with the company from approximately 10.5 years to 10.8 years. All of these times are unusually long.

i) Since the data is normal, the low outlier cutoff is the mean − (2 x standard deviation) = 7.345 − (2 x 1.376) = 4.593. So any employees that have been with the company 4.593 years or less is considered an unusually small amount of time. The dot plot shows five employees that have been with the company from approximately 3.6 years to 4.5 years. All of these times are unusually short.

17.

a) Driving alone was most popular.

b) Biking was least popular.

c) 10 of the stat students walked to school.

d) 0.054

e) 0.018 x 100% = 1.8%

1.8% of the stat students used public transportation.

18.

a) 34%

b) 16%

c) Typical salaries are between 29.3 thousand dollars and 33.5 thousand dollars.

d) Salaries above 35.6 thousand dollars are considered unusually high (high outliers).

e) Salaries below 27.2 thousand dollars are considered unusually low (low outliers).

f) The average salary is 31.4 thousand dollars.
