

### Section 1A

1. Bear Ages: Quantitative, Months  
Bear Month Data Taken: Categorical, 8 options (May-November)  
Bear Gender: Categorical, 2 options  
Head Length: Quantitative, Inches  
Head Width: Quantitative, Inches  
Neck Circumference: Quantitative, Inches  
Length: Quantitative, Inches  
Chest: Quantitative, Inches  
Weight: Quantitative, Pounds
  
  3.
    - a) Milligrams of Aspirin: Quantitative
    - b) Types of Cars: Categorical
    - c) Smoke Marijuana or not: Categorical
    - d) Number of Bicycles: Quantitative
    - e) Types of Birds: Categorical
    - f) Grams of Gold: Quantitative
    - g) Types of Cardio Classes: Categorical
    - h) Number of Cardio Classes: Quantitative
    - i) City: Categorical
    - j) Money in Bank Accounts: Quantitative
    - k) Zip Codes: Categorical
    - l) Driver's License Numbers: Categorical
    - m) Number of Taxis: Quantitative
- 

### Section 1B

1. Population of Interest: All students at the college.  
Method: Voluntary Response  
Will not represent the population very well. There is sampling Bias, since the individuals were not chosen randomly.
  
3. Population of Interest: All students at the high school.  
Method: Convenience  
Will not represent the population very well. There is sampling Bias, since the individuals were not chosen randomly.
  
5. Population of Interest: All people in Rachael's home town.  
Method: Systematic  
Might represent the population. There is sampling bias, since the individuals were not chosen randomly, but it may be representative since the whole population was on the list. This data is not as biased as convenience or voluntary response, but not as good as a random sample.
  
7. Population of Interest: All employees at the company.  
Method: Census  
Census is better than a random sample. Will represent the population very well as long as there is no other types of bias present. No sampling bias.
  
9. Population of Interest: All people in Toronto.  
Method: Simple Random Sample  
Will represent the population well as long as there is no other types of bias present. No sampling bias.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

11. Population of Interest: All people that use smart phones.  
Method: Voluntary Response  
Will not represent the population very well. Sampling bias, since the individuals were not chosen randomly.
13. Population of Interest: All teenagers and adults.  
Method: Stratified since they are comparing groups.  
Will represent the population well as long as there is no other types of bias present. Individuals were chosen randomly, so no sampling bias.
15. Population of Interest: All adults in North Carolina  
Method: Systematic and Convenience  
Will not represent the population very well. There is sampling bias since the individuals were not selected randomly. This data is particularly bad since most of the population has no opportunity to enter the store.
- 

### Section 1C

- 1.
- a) Population: All people or objects to be studied. For example, all students at College of the Canyons.
  - b) Census: Collecting data from everyone in your population. For example, collecting data from all of the students at college of the canyons.
  - c) Sample: Collecting data from a subgroup of the population. For example, collecting data from fifty students at College of the Canyons.
  - d) Bias: When data does not reflect the population. For example, friends and family will not represent the population of all people in Los Angeles, CA.
  - e) Question Bias: Phrasing a question in order to force people to answer the way you want. For example, we want to collect data on smoking cigarettes, but give the person a lecture on how unhealthy cigarettes are before asking them.
  - f) Response Bias: When someone is likely to lie about the answer to a question. For example, asking people how much they weigh in pounds. They may not give you a truthful answer.
  - g) Sampling Bias: Not using randomization when collecting sample data. For example, collecting data from only your friends and family. This is not a random sample.
  - h) Deliberate Bias: Falsifying or changing your data or leaving out groups from your population of interest. For example, a person might remove all of the data from people that disagreed with their opinion.
  - i) Non-response Bias: When people are likely to not answer when asked to provide data. Randomly calling phone numbers to get data, but the person refuses to answer the phone.

3. Population of interest: All people in the U.S.

Question Bias: The question was phrased to make people feel bad about answering no.

Response Bias: Vaccinations are a controversial issue and many people may feel scared to admit that they don't agree with vaccinations.

Non-response Bias: There will be many people that randomly selected, but refuse to answer the question.

5. Population of interest: All Americans.

Response Bias: Cocaine users would not feel comfortable answering the question honestly.

Non-response: Many people may be randomly selected, but will chose not to answer the question.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

7. Population of interest: All adults in Palmdale, CA.

Sampling Bias: The individuals were not selected randomly.

Deliberate Bias: Julie skipped streets that looked poor. These people are not being represented in the data.

Response Bias: People often lie about their income.

Non-response: Many people may not be home or refuse to answer the door.

9. Population of interest: All pills made by the company.

Deliberate Bias: They deleted data that poorly reflected the pharmaceutical company.

---

## Section 1D

1. Explanatory Variable: Having a cell phone or not.

Response Variable: Ruler catch length in inches or “drop”.

2. We needed a control group that measures the classes ability to catch a ruler in general. We can then compare the cell phone data to the control group.

3. The two groups were people with the cell phone (treatment group) and those without a cell phone (control group). They were perfectly alike in all confounding variables since they were the same exact people measured twice.

4. Answers may vary. Confounding Variables: Age, hand-eye coordination, distractions besides the phone, hand size, ability to text one handed, position of the ruler when dropped, ...

Since the same people were measured twice, the two groups had the exact same ages, hand-eye coordination and ability to text one-handed. The amount of distraction was relatively the same with or without the phone. The instructor gave a demonstration so that everyone would hold and drop the ruler the same way.

5. Neither. The explanatory variable was having a cell phone or not. The person knew whether they had a cell phone or not. Not knowing when the ruler would be dropped does not constitute blind since it is the response variable.

6. Answers will vary from class to class. The average catch distance was lower for the no cell phone group, but it is difficult to determine if it is significant at this point. We will learn that later. The number of drops was significantly greater in the cell phone group. Since confounding variables were controlled, we have proven that texting does cause you to drop the ruler more often. Whether this experiment applies to texting while driving is debatable. Some people have said that dropping the ruler may be equivalent to not hitting the breaks in time. Again, this is debatable. The experiment does prove that texting slows reflexes and you do need reflexes when you drive.

7. Observational Study: Collecting data without trying to control confounding variables. Data collected by an observational study can show relationships but cannot prove cause and effect.

8. Experiment: A scientific method for controlling confounding variables and proving cause and effect.

9. Explanatory Variable: The independent or treatment variable. In an experiment, this is the variable that causes the effect.

10. Response Variable: The dependent variable. In an experiment this is the variable that measures the effect.

11. Confounding Variables (or lurking variables): Other variables that might influence the response variable other than the explanatory variable being studied.

12. 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.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

13. Placebo: A fake medicine or fake treatment used to control the placebo effect.
14. Placebo Effect: The capacity of the human brain to manifest physical responses based on the person believing something is true.
15. Single Blind: When only the person receiving the treatment does not know if it is real or a placebo.
16. Double Blind: When both the person receiving the treatment and the person giving the treatment does not know if it is real or a placebo.
17. This is an experiment, since they need to prove that the medicine has the desired effect. They must control confounding variables like the amount of motion, genetics, age, diet, pregnancy, etc. If they control all of the confounding variables and the medicine (treatment) group has significantly less motion sickness, they will have succeeded in proving cause and effect.
19. This is an observational study since they just collected data without thought to controlling confounding variables. This can show the number of cases of tuberculosis is related to or associated with the low income, crowded cities, but it will not be able to prove cause and effect.
21. This is an observational study since they just collected data without thought to controlling confounding variables. This can show that obesity is related to or associated with having diabetes, but they will not be able to prove cause and effect. There are many variables involved in determining why someone has diabetes.
- 

**Section 1E**

- 1.
- a) 0.75
  - c) 0.00664
  - e) 0.397
  - g) 0.00189
  - i) 0.0316
  - k) 0.961
  - m) 0.00007
  - o) 0.662
  - q) 1
- 2.
- a) 5.7%
  - c) 0.33%
  - e) 6.13%
  - g) 0.045%
  - i) 4.6%
  - k) 0.27%
  - m) 0.58%
  - o) 100%
  - q) 2.04%



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

3.

$$15\% = 0.15$$

$$\text{Estimated Amount} = 0.15 \times 78300 = 11,745$$

We estimate that approximately 11,745 people in Chino Hills are without health insurance.

5.

$$9.3\% = 0.093$$

$$\text{Estimated Amount} = 0.093 \times 18400 = 1711.2 \approx 1711$$

We estimate that approximately 1,711 students at COC have diabetes.

7.

$$1.47\% = 0.0147$$

$$\text{Estimated Amount} = 0.0147 \times 136400 = 2005.08 \approx 2005$$

We estimate that approximately 2,005 people in Van Nuys have autism.

9.

$$14.8\% = 0.148$$

$$\text{Estimated Amount} = 0.148 \times 305700 = 45243.6 \approx 45244$$

We estimate that approximately 45,244 people in Stockton live below the poverty line.

11.

$$\text{Athletic Wear: } 139/213 \approx 0.653 = 65.3\%$$

$$\text{Traditional Jeans: } 74/213 \approx 0.347 = 34.7\%$$

$$\text{Percent of Increase} = (0.653 - 0.347)/0.347 = 0.882 = 88.2\% \text{ of increase.}$$

The percent of women that prefer athletic wear does seem to be significantly higher than the percent that prefer traditional jeans. It is also practically significant since there was 65 more women in the sample that preferred athletic wear.

13.

$$\text{Med/Surg: } 57/350 \approx 0.163 = 16.3\%$$

$$\text{Telemetry: } 49/350 \approx 0.14 = 14\%$$

$$\text{Percent of Increase} = (0.163 - 0.14)/0.14 \approx 0.164 = 16.4\% \text{ of increase.}$$

The percent of patients admitted to telemetry and med/surge seem very close. The percent of increase is very small. There were also only 8 more patients in Med/Surg than telemetry. These indicate there is no significant difference, practically or statistically.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

15.

Medicine:  $13/57 \approx 0.228 = 22.8\%$

Placebo:  $11/61 \approx 0.180 = 18.0\%$

Percent of Increase =  $(0.228 - 0.18)/0.18 \approx 0.267 = 26.7\%$  of increase.

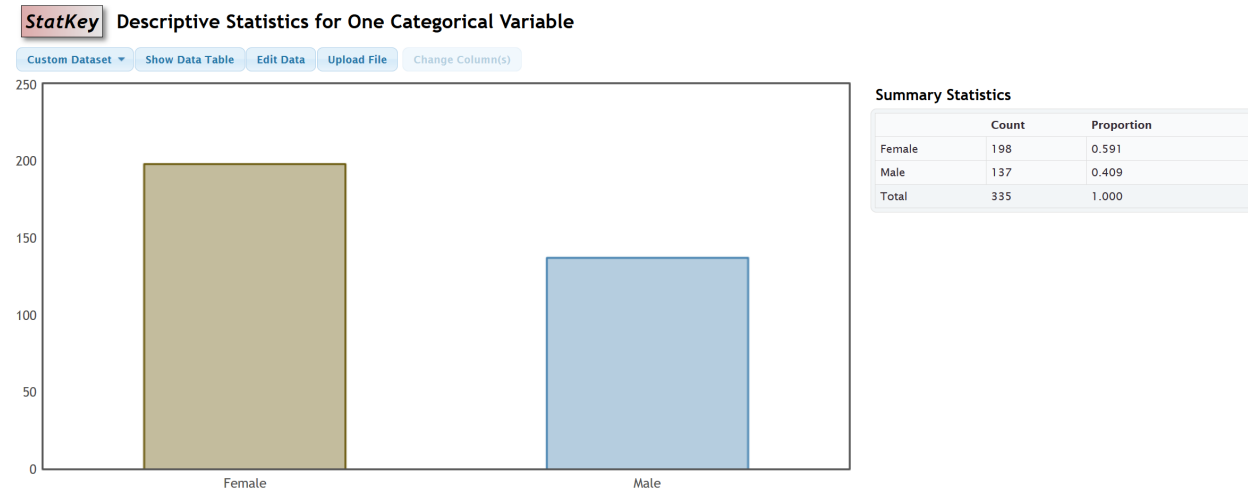
The percent of patients that improved on the medicine is only slightly higher than the placebo group. Practically there is not much difference. Only two more patients on the medicine showed improvement. That is not practically significant.

17.

Proportion of female  $\approx 0.591$

Proportion of male  $\approx 0.409$

Percent of Increase =  $(0.591 - 0.409)/0.409 \approx 0.445 = 44.5\%$  of increase. This indicates that the percentage of female COC statistics students is significantly higher than the percentage of male students. It is also practically significant since there were 61 more female students than male.



19.

Proportion of Democrat  $\approx 0.335$

Proportion of Republican  $\approx 0.192$

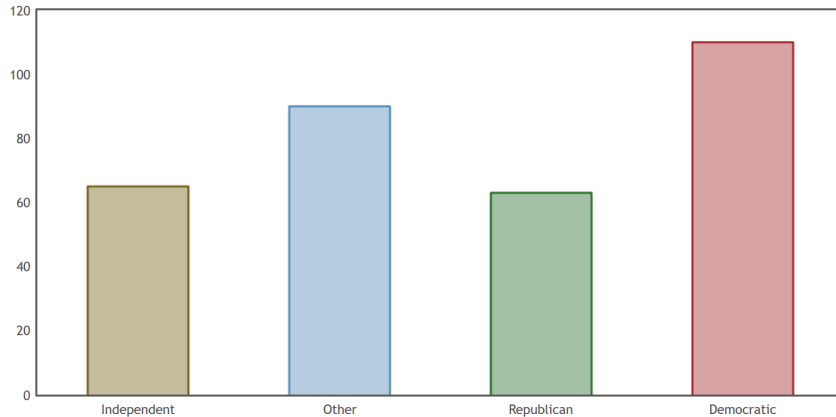
Percent of Increase =  $(0.335 - 0.192)/0.192 \approx 0.745 = 74.5\%$  of increase. This indicates that the percentage of COC statistics students that are democrat is significantly higher than republican. It is practically significant also since there are 47 more democratic statistics students than republican.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

StatKey Descriptive Statistics for One Categorical Variable

Custom Dataset Show Data Table Edit Data Upload File Change Column(s)



Summary Statistics

	Count	Proportion
Independent	65	0.198
Other	90	0.274
Republican	63	0.192
Democratic	110	0.335
Total	328	1.000

21.

Percentage of cars made in France  $\approx 3\%$

Number of cars made in U.S. = 22

Proportion of cars made in Sweden  $\approx 0.05$

Proportion of cars made in Japan  $\approx 0.18$

Proportion of cars made in Germany  $\approx 0.13$

Percent of Increase (Japan & Germany)  $\approx (0.18 - 0.13) / 0.13 \approx 0.385 = 38.5\%$

The percent of increase does seem to indicate that the percent of cars made in Japan is significantly higher than Germany. However does not seem to be practically significant since the number of cars made in Japan was only 2 more than Germany. Overall, I would say it is not significant.

23.

Percentage of cereals made by Quaker  $\approx 17\%$

Number of cereals made by Ralston = 3

Proportion of cereals made by General  $\approx 0.29$

Proportion of cereals made by Kelloggs  $\approx 0.33$

Proportion of cereals made by Quaker  $\approx 0.17$

Percent of Increase (Kelloggs & Quaker)  $\approx (0.33 - 0.17) / 0.17 \approx 0.941 = 94.1\%$

The percent of increase indicates that the percent of cereals made by Kelloggs is significantly higher than Quaker. However does not seem to be practically significant since it was a small sample size and the number of cereals made by Kelloggs was only 4 more than Quaker.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

25.

Percentage of cereals on top shelf  $\approx 33\%$

Number of cereals on bottom shelf = 8

Proportion of cereals on middle shelf  $\approx 0.33$

Percent of Increase (top and bottom shelves)  $\approx (0.33-0.33) / 0.33 \approx 0 = 0\%$

There is no significant difference between the percentages of cereals put on the top and bottom shelves. They appear to be about the same.

27.

a)  $0.122 = 12.2\%$

b)  $0.113 = 11.3\%$

c)  $0.796 = 79.6\%$

d)  $1 - 0.796 = 0.204 = 20.4\%$

e)  $0.110 = 11.0\%$

f)  $1 - 0.110 = 0.89 = 89\%$

**Binomial Distribution:  $n=84$ ,  $p=0.12$**

Input: 11.0

Type: Probability density

X P(X)

11.0 0.122219

**Binomial Distribution:  $n=84$ ,  $p=0.12$**

Input: 8.0

Type: Probability density

X P(X)

8.0 0.113128

**Binomial Distribution:  $n=84$ ,  $p=0.12$**

Input: 12.0

Type: Cumulative probability

X P( $\leq$ X)

12.0 0.796146

**Binomial Distribution:  $n=84$ ,  $p=0.12$**

Input: 6.0

Type: Cumulative probability

X P( $\leq$ X)

6.0 0.109721





Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

29.

a)  $0 = 0\%$

b)  $1 - 0 = 1 = 100\%$

**Binomial Distribution:  $n=57$ ,  $p=0.845$**

Input: 9.0

Type: Cumulative probability

X P( $\leq X$ )

9.0 0

---

### Section 1F

1.

a) The shape of the data is relatively bell shaped or unimodal and symmetric. In a histogram, the tallest bars are relatively in the middle and the left and right tail are about the same length.

b) The mean average is the average that we use when the data is normal. It also balances the distances. It is calculated by adding all of the data values and dividing the sum by the sample size  $n$ .

c) The standard deviation calculates the average distance that data values are from the mean. It is the measure of spread used when data is normal. To calculate the standard deviation for one quantitative data set, take every number in the data set and subtract the mean from it. Then square are the differences. Add up all the squares and divide by  $n - 1$  where  $n$  is the sample size. Now take the square root of the answer. Always have a computer calculate the standard deviation for you.

2.

a) Mean Average

b) Standard Deviation

c) One Standard Deviation or less.

d) Mean  $\pm$  Standard Deviation

e) 68%

f) Two or more standard deviations.

g) 2.5%

h) 2.5%

3.

a) The data measures the neck circumference of bears. The units are inches.

b) 54 total bears

c) Yes. The data is nearly normal (almost bell shaped).

d) 10 inches



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

e) 31.5 inches

f) Center = Mean Average = 20.556 inches

g) Typical Spread = Standard Deviation = 5.641 inches

h)

$$20.556 - 5.641 = 14.915 \text{ inches}$$

$$20.556 + 5.641 = 26.197 \text{ inches}$$

Typical bears have a neck circumference between 14.915 inches and 26.197 inches.

i)

$$\text{Unusual High Cutoff: } 20.556 + (2 \times 5.641) = 20.556 + 11.282 = 31.838 \text{ inches.}$$

Any neck circumference of 31.838 inches or more would be considered unusually high (high outlier).

j)

$$\text{Unusual Low Cutoff: } 20.556 - (2 \times 5.641) = 20.556 - 11.282 = 9.274 \text{ inches.}$$

Any neck circumference of 9.274 inches or less would be considered unusually low (low outlier).

k)

There are no unusually large bear neck sizes. The largest neck size is 31.5 inches which is not higher than the unusually high cutoff of 31.838 inches.

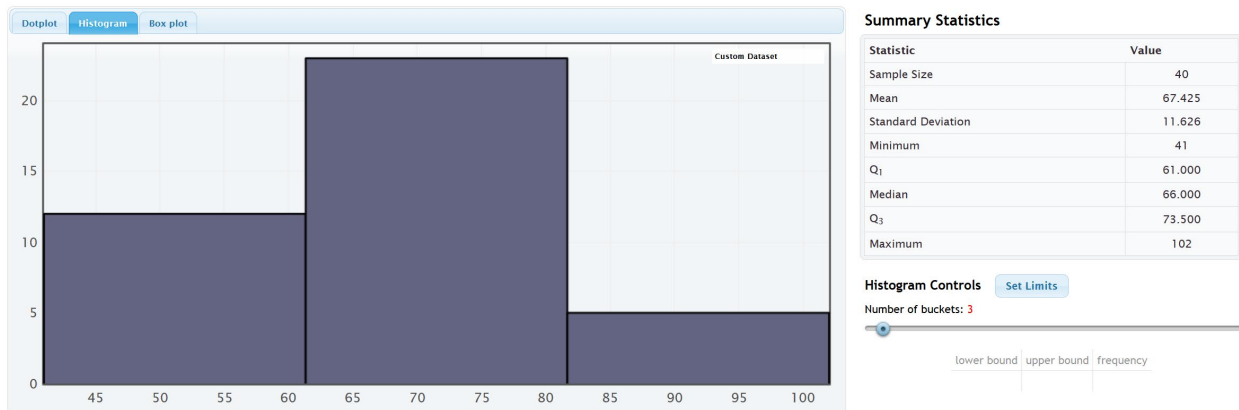
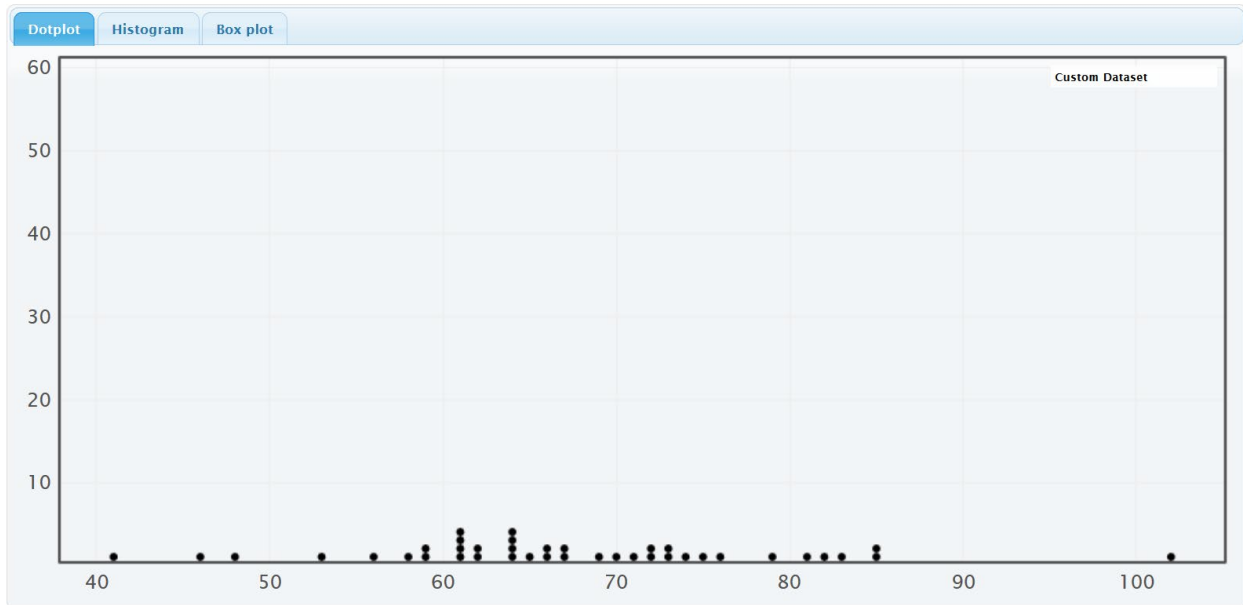
l)

There are no unusually small bear neck sizes. The smallest neck size is 10 inches which is not lower than the unusually low cutoff of 9.274 inches.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

5.



- The data is measuring the diastolic blood pressure of women. The units are millimeters of mercury (mm of Hg).
- There were 40 total women in the data set.
- The data is nearly normal (almost bell shaped).
- Min = 41 mm of Hg
- Max = 102 mm of Hg
- Center = Mean Average = 67.425 mm of Hg
- Typical Spread = Standard Deviation = 11.626 mm of Hg
- 

Mean – Standard Deviation = 67.425 – 11.626 = 55.799 mm of Hg



This chapter is from *Introduction to Statistics for Community College Students, 1<sup>st</sup> Edition*, by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed under a "CC-By" [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) – 10/1/18

Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

Mean + Standard Deviation =  $67.425 + 11.626 = 79.051$  mm of Hg

Typical women in this data have a diastolic blood pressure between 79.051 mm of Hg and 55.799 mm of Hg.

i)

Unusual High Cutoff:  $67.425 + (2 \times 11.626) = 67.425 + 23.252 = 90.677$  mm of Hg.

Any woman in the data with a diastolic blood pressure of 90.677 mm of Hg or higher would be considered unusually high (high outlier).

j)

Unusual Low Cutoff:  $67.425 - (2 \times 11.626) = 67.425 - 23.252 = 44.173$  mm of Hg.

Any woman in the data with a diastolic blood pressure of 44.173 mm of Hg or lower would be considered unusually low (low outlier).

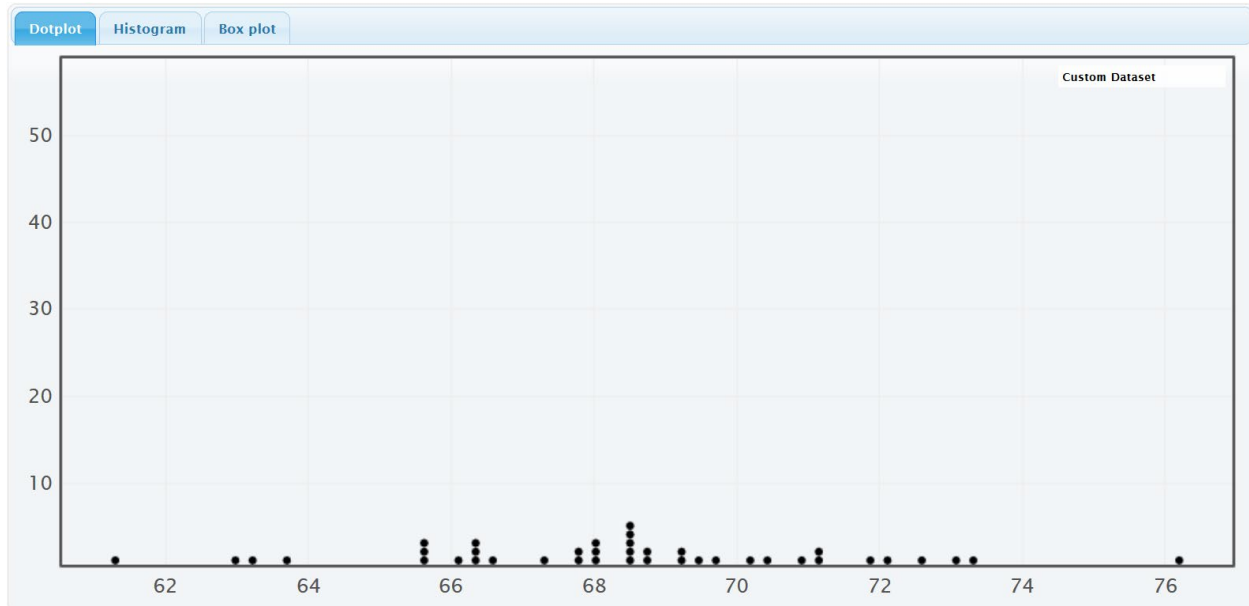
k)

The dot plot shows only one dot above the unusual high cutoff of 90.677 mm of Hg. It is the maximum value of 102 mm of Hg. So the woman with a diastolic blood pressure of 102 mm of Hg is considered unusually high or a high outlier.

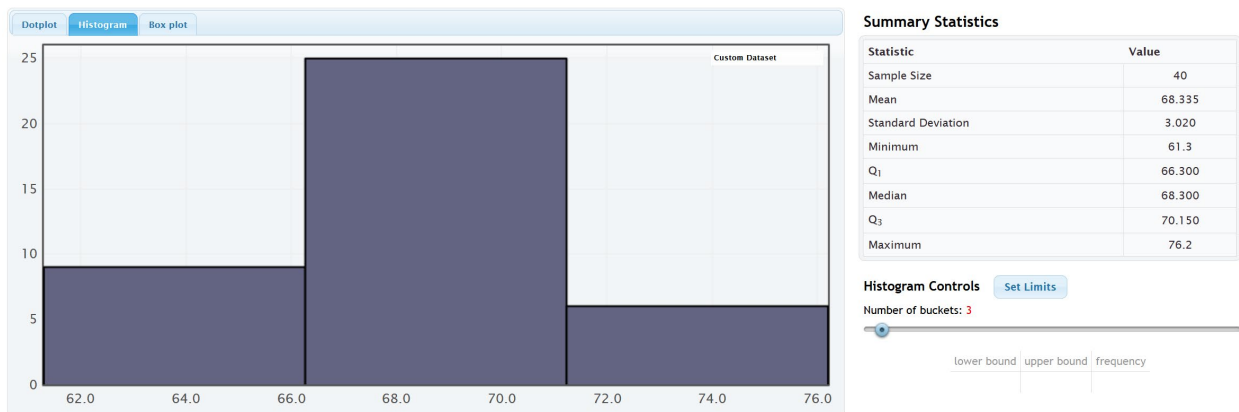
l)

The dot plot shows only one dot below the unusual low cutoff of 44.173 mm of Hg. It is the minimum value of 41 mm of Hg. So the woman with a diastolic blood pressure of 41 mm of Hg is considered unusually low or a low outlier.

7.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1



a) The data is measuring the heights of men. The units are inches.

b) There were 40 total men in the data set.

c) The data is nearly normal (almost bell shaped).

d) Min = 61.3 inches

e) Max = 76.2 inches

f) Center = Mean Average = 68.335 inches

g) Typical Spread = Standard Deviation = 3.020 inches

h)

Mean – Standard Deviation =  $68.335 - 3.020 = 65.315$  inches

Mean + Standard Deviation =  $68.335 + 3.020 = 71.355$  inches

Typical men in this data have a height between 65.315 inches and 71.355 inches.

i)

Unusual High Cutoff:  $68.335 + (2 \times 3.020) = 68.335 + 6.040 = 74.375$  inches.

Any man in the data with a height of 74.375 inches or more would be considered unusually high (high outlier).

j)

Unusual High Cutoff:  $68.335 - (2 \times 3.020) = 68.335 - 6.040 = 62.295$  inches.

Any man in the data with a height of 62.295 inches or less would be considered unusually low (low outlier).

k)

The dot plot shows only one dot above the unusual high cutoff of 74.375 inches. It is the maximum value of 76.2 inches. So the man with a height of 76.2 inches is considered unusually tall or a high outlier.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

l)

The dot plot shows only one dot below the unusual low cutoff of 62.295 inches. It is the minimum value of 61.3 inches. The next largest dot was 62.9 inches which is not below the cutoff. So the man with a height of 61.3 inches is considered unusually short or a low outlier.

9.

A Z-score is the number of standard deviations that a value is from the mean.

10.

Typical values have a Z-score between  $-1$  and  $+1$  inclusively.

11.

For normal data, a Z-score of  $+2$  or higher would indicate that the data value is unusually high or a high outlier.

For normal data, a Z-score of  $-2$  or less would indicate that the data value is unusually low or a low outlier.

13.

a)  $Z = (89 - 99.8) / 15.3 \approx -0.71$

b) Jan's IQ score was only 0.71 standard deviations below the mean.

c) Jan's IQ is not unusual, since the Z-score is not  $+2$  or above or  $-2$  or below. In fact, she has a very typical IQ, since the Z-score since it is between  $-1$  and  $+1$ .

15.

a)  $Z = (13.61 - 46.89) / 12.44 \approx -2.68$

b) The amount of money that Julie spent is 2.68 standard deviations below the mean.

c) The amount that Julie spent was unusually low (low outlier) since the Z-score was below  $-2$ .

17.

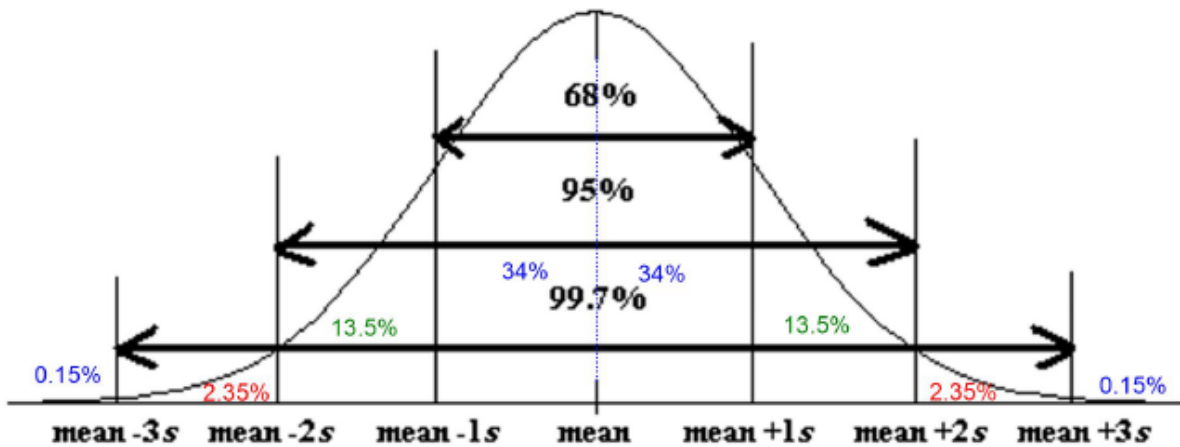
a)  $Z = (57 - 35.663) / 9.352 \approx +2.28$

b) This bears chest size is 2.28 standard deviations above the mean.

c) This bears chest size is unusually large (high outlier) since the Z-score is greater than  $+2$ .



19.



21.

a)  $34\% + 34\% + 13.5\% = 81.5\%$

b)  $34\% + 34\% + 13.5\% + 2.35\% + 0.15\% = 84\%$

c) One standard deviation from the mean is typical. So typical bear neck circumferences are between 14.915 inches and 26.197 inches.

d) The unusual high cutoff is two standard deviations above the mean which is 31.838 inches. So any bear with a neck circumference of 31.838 inches or more is considered unusually large.

e) The unusual low cutoff is two standard deviations below the mean which is 9.274 inches. So any bear with a neck circumference of 9.274 inches or less is considered unusually small.

f) One standard deviation below the mean has 84% above. So 84% of the bears have a neck circumference greater than 14.915 inches.

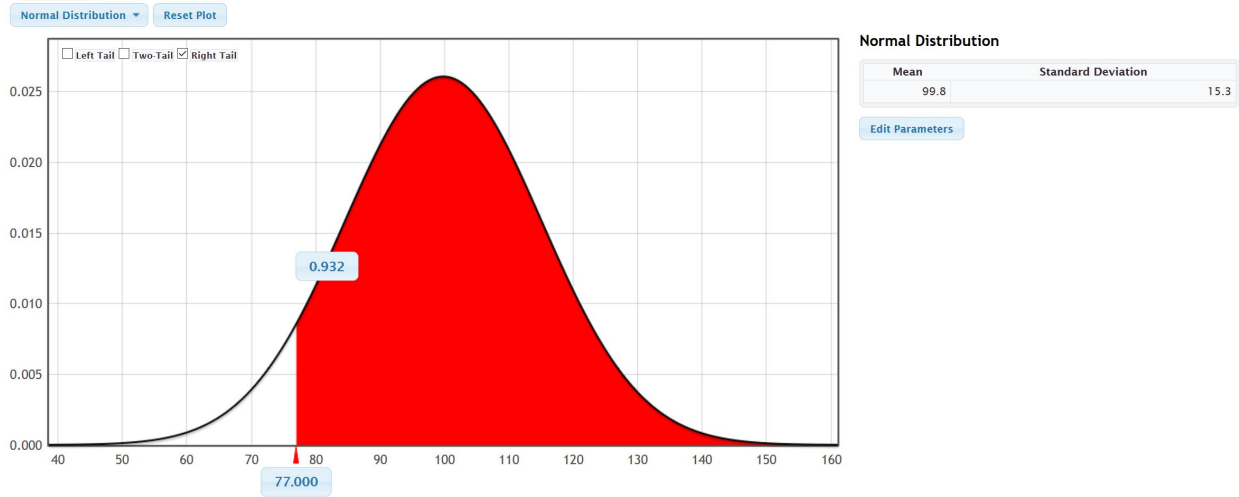
g)  $13.5\% + 2.35\% + 0.15\% = 16\%$



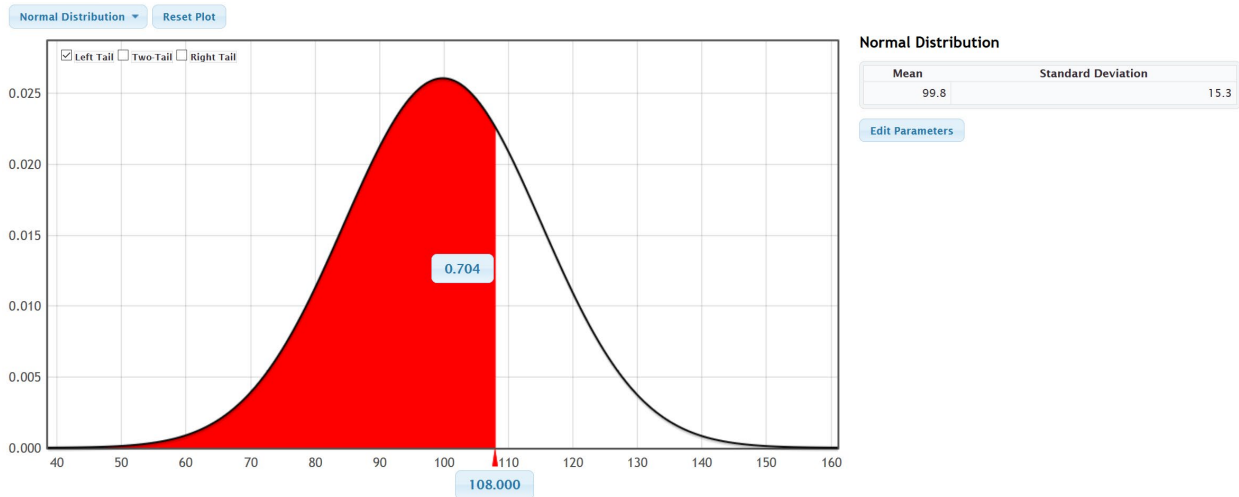
Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

23.

a) Based on this data, about 93.2% of people have an IQ above 77.



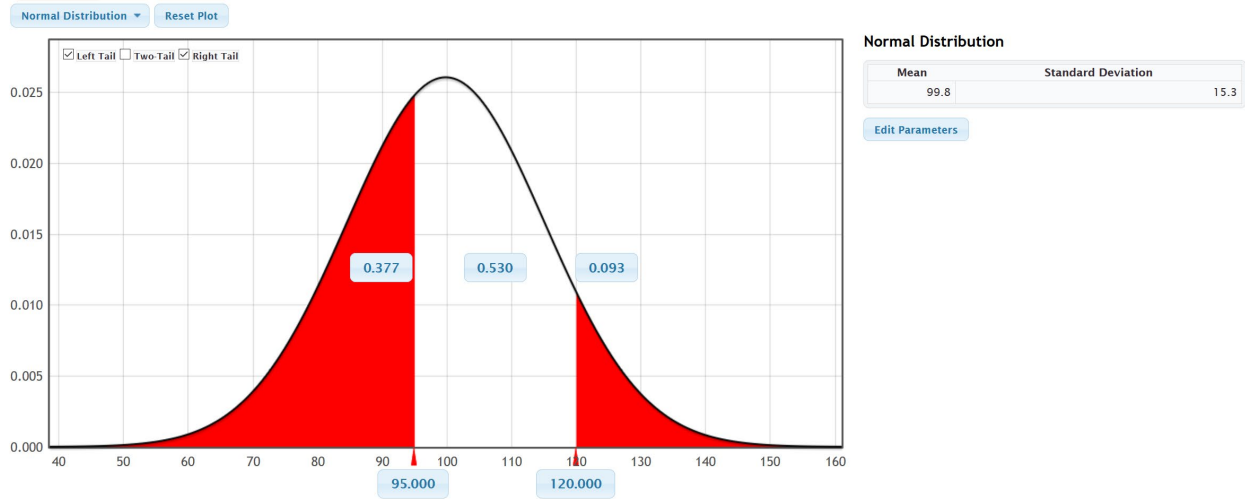
b) Based on this data, about 70.4% of people have an IQ below 108.



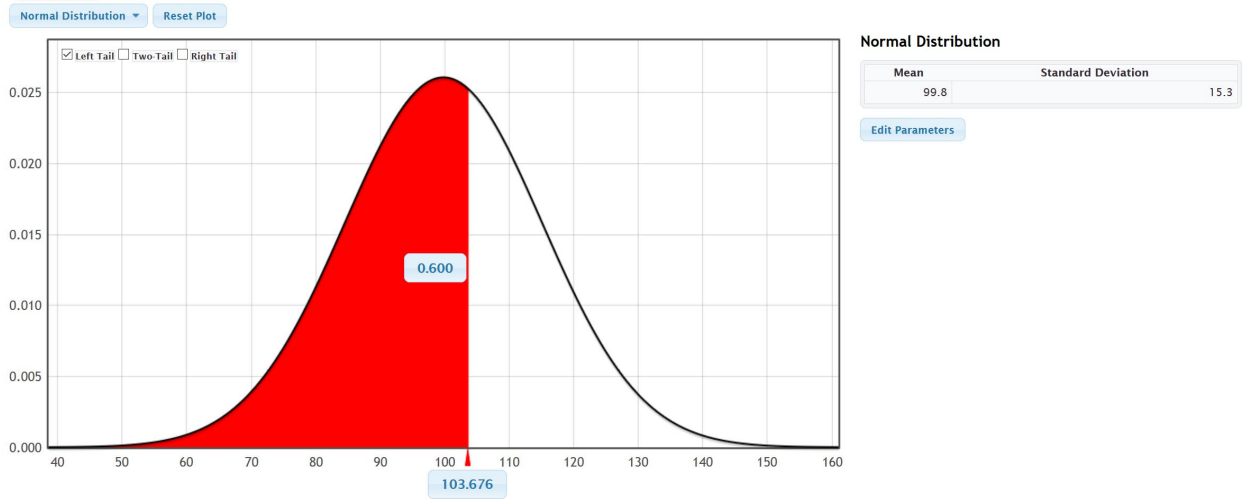


Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

c) Based on this data, about 53.0% of people have an IQ between 95 and 120.

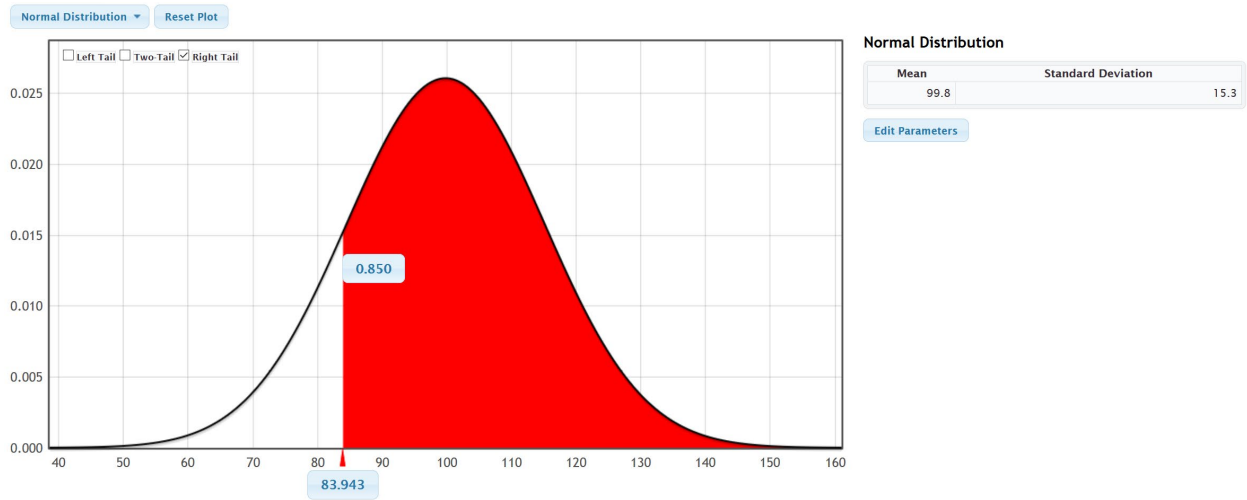


d) About 60% of people have an IQ below 103.676.

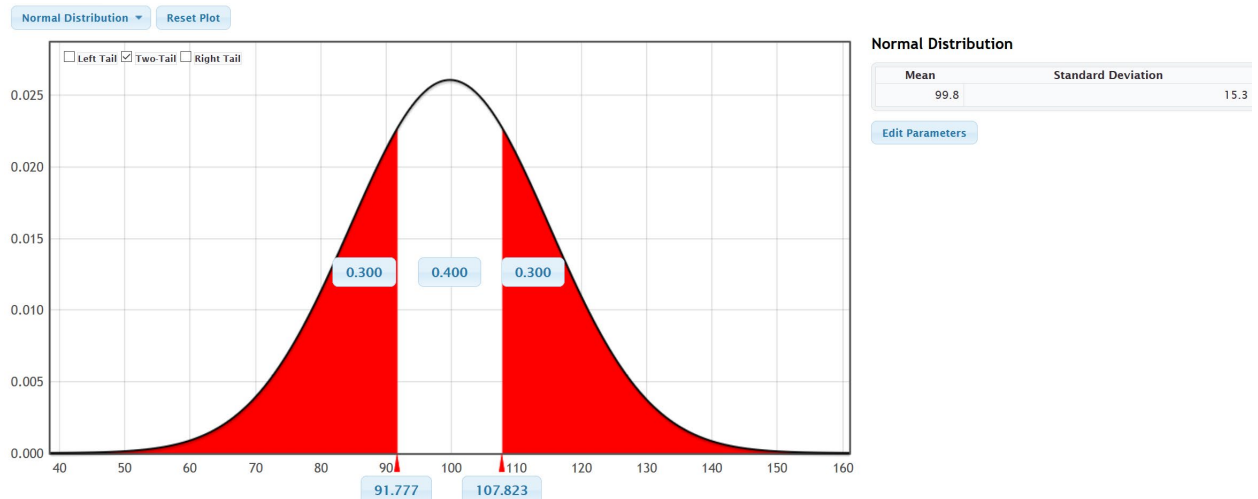


Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

e) Based on this data, about 85% of people have an IQ above 83.943.



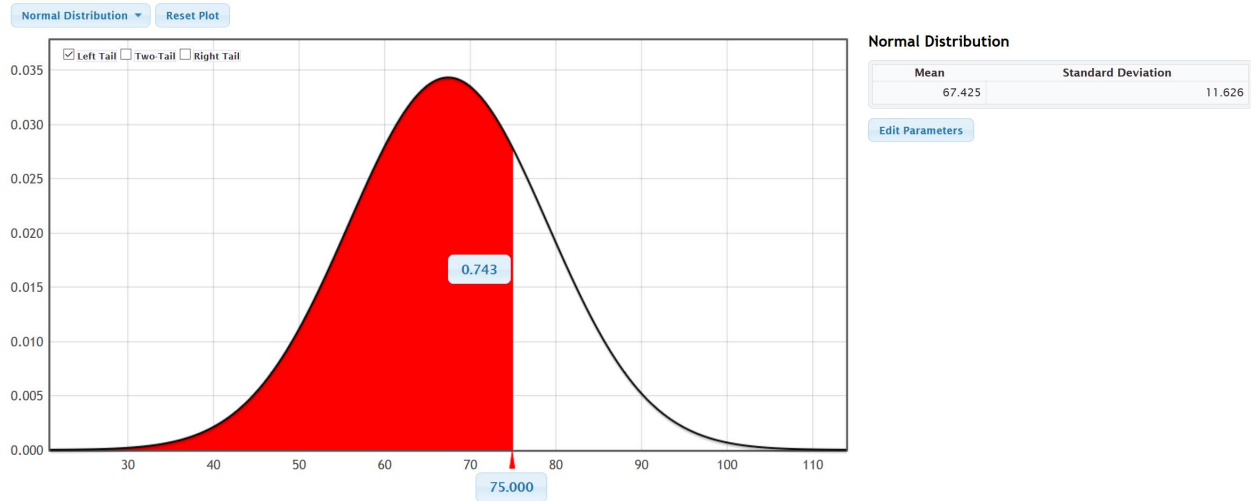
f) Based on this data, the middle 40% of people have an IQ between 91.777 and 107.823.



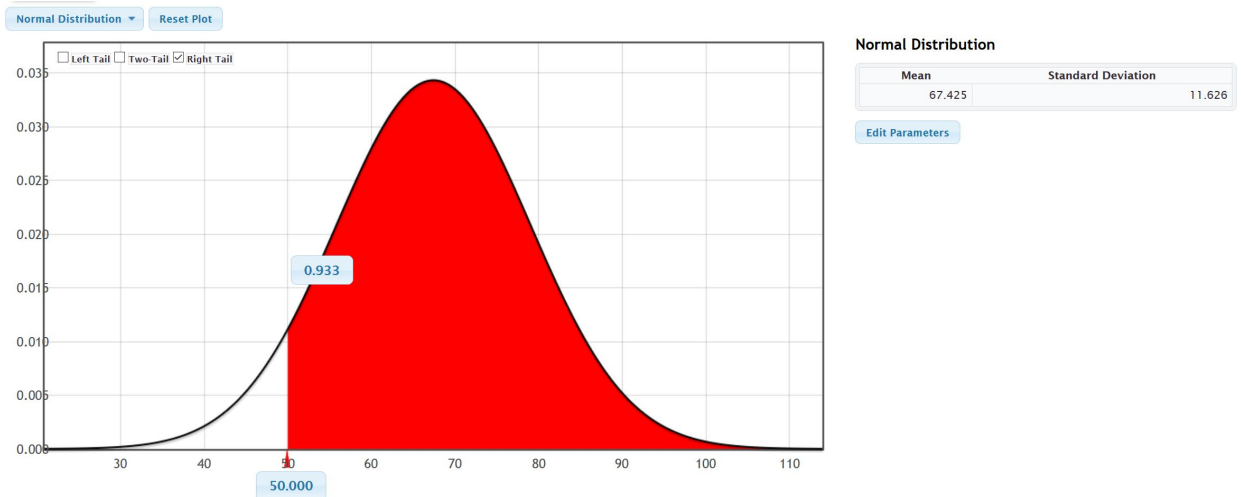
Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

25.

a) Based on this data, about 74.3% of women have a diastolic blood pressure below 75 mm of Hg.



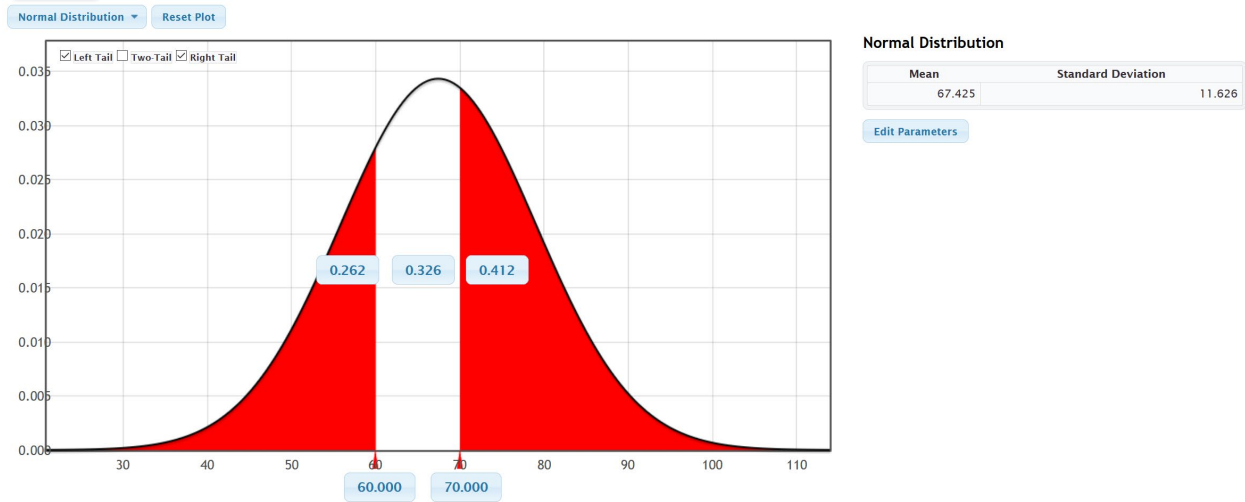
b) Based on this data, about 93.3% of women have a diastolic blood pressure above 50.



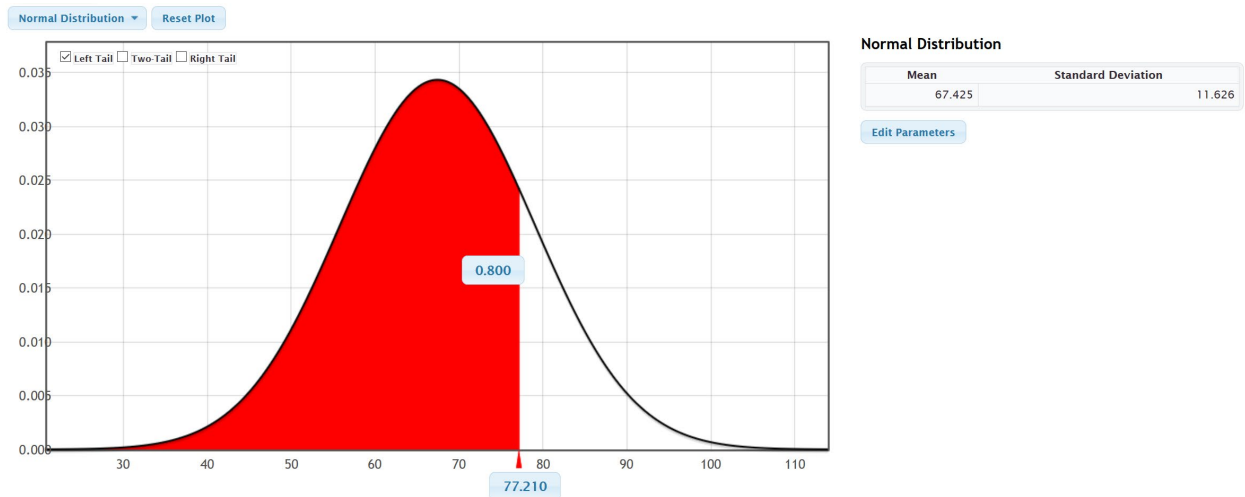
This chapter is from *Introduction to Statistics for Community College Students, 1<sup>st</sup> Edition*, by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed under a "CC-By" [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) – 10/1/18

Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

c) Based on this data, about 32.6% of women have a diastolic blood pressure between 60 and 70.

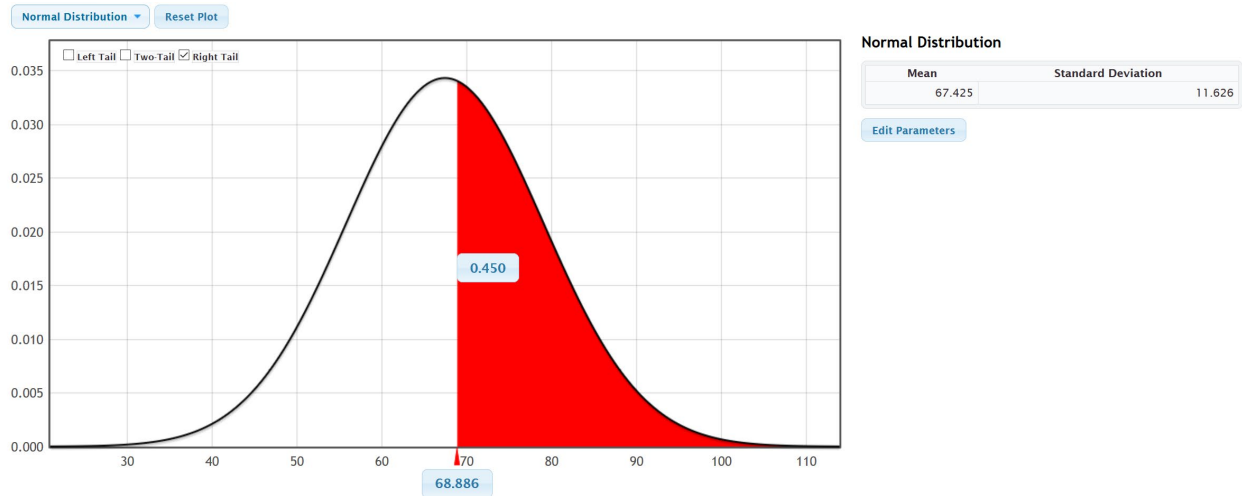


d) Based on this data, about 80% of women have a diastolic blood pressure less than 77.21 mm of Hg.

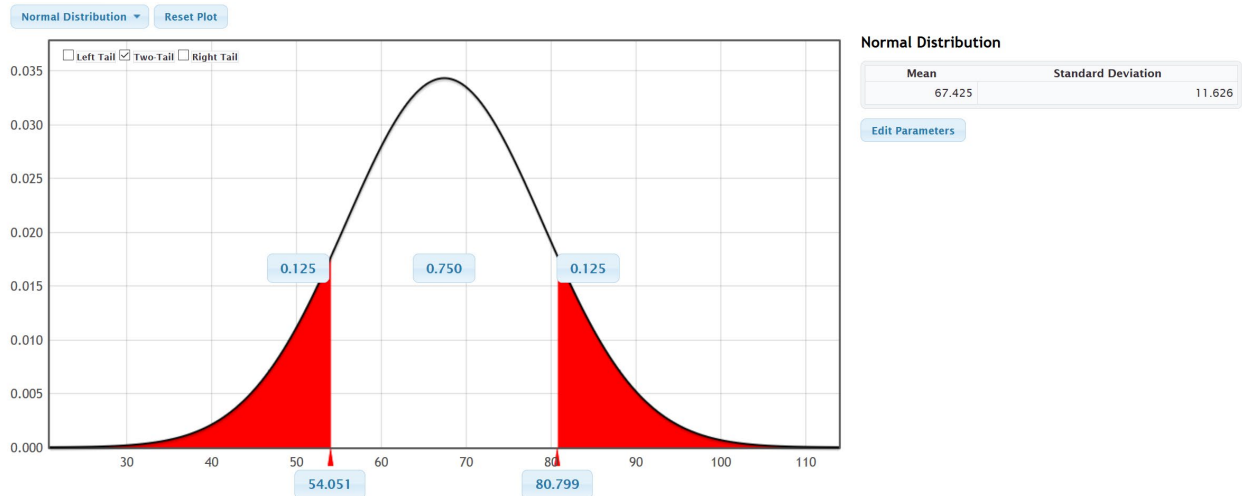


Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

e) Based on this data, about 45% of women have a diastolic blood pressure above 68.886 mm of Hg.



f) Based on this data, the middle 75% of women's diastolic blood pressures fall between 54.051 mm of Hg and 80.799 mm of Hg.



## Section 1G

1.

- a) A skewed right shape has the center on the far left and a long tail to the right. A histogram would have the highest bars on the far left with a short left tail and a long right tail.
- b) A skewed left shape has the center on the far right and a long tail to the left. A histogram would have the highest bars on the far right with a short right tail and a long left tail.
- c) The median average is the average or center when the data values are put in order. When data sets are not normal, we prefer to use the median as our average. The median also splits the data so that approximately 50% of the data is above the median and 50% of the data is below the median. To calculate the median, first put the numbers in order. If there is one number in the middle, then that is the median. If there are two numbers in the middle, then the median will be half way between the two numbers in the middle.
- d) The first quartile is the number that approximately 25% of the data values are less than and 75% of the data values are greater than. To calculate the first quartile, simply calculate the median of the bottom half of the data when the data values are in order.
- e) The third quartile is the number that approximately 75% of the data values are less than and 25% of the data values are greater than. To calculate the third quartile, simply calculate the median of the top half of the data when the data values are in order.
- f) The interquartile range is the best measure of typical spread for non-normal data. It measures the distance between the middle 50% of the data values. It can also be thought of as the maximum distance between typical values in a non-normal data set. To calculate IQR, subtract the third quartile minus the first quartile.

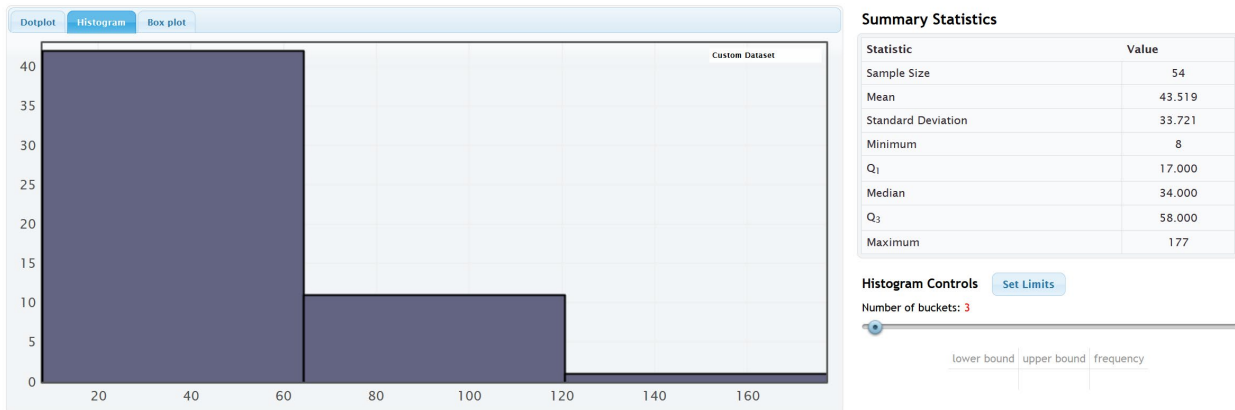
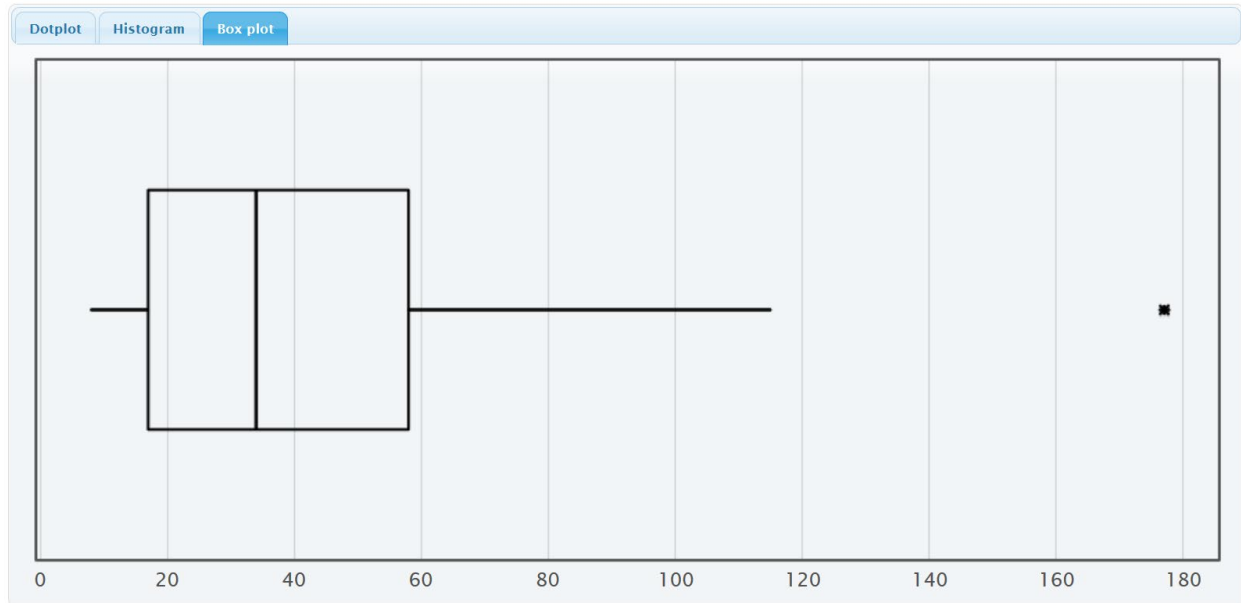
2.

- a) If the data is not normal, we should use the median as our average or center.
- b) If the data is not normal, we should use the IQR as the best measure of typical spread.
- c) If the data is not normal, then typical values will fall between the 1<sup>st</sup> quartile and the 3<sup>rd</sup> quartile.
- d) The middle 50% is typical for data that is not normal.
- e) If the data is not normal, then you can use the box plot to identify unusually high values (high outliers). For horizontal box plots, look for circles, triangles or stars to the far right of the right whisker. For vertical box plots, look for circles, triangles or stars above the top whisker.
- f) If the data is not normal, then you can use the box plot to identify unusually low values (low outliers). For horizontal box plots, look for circles, triangles or stars to the far left of the left whisker. For vertical box plots, look for circles, triangles or stars below the bottom whisker.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

3.



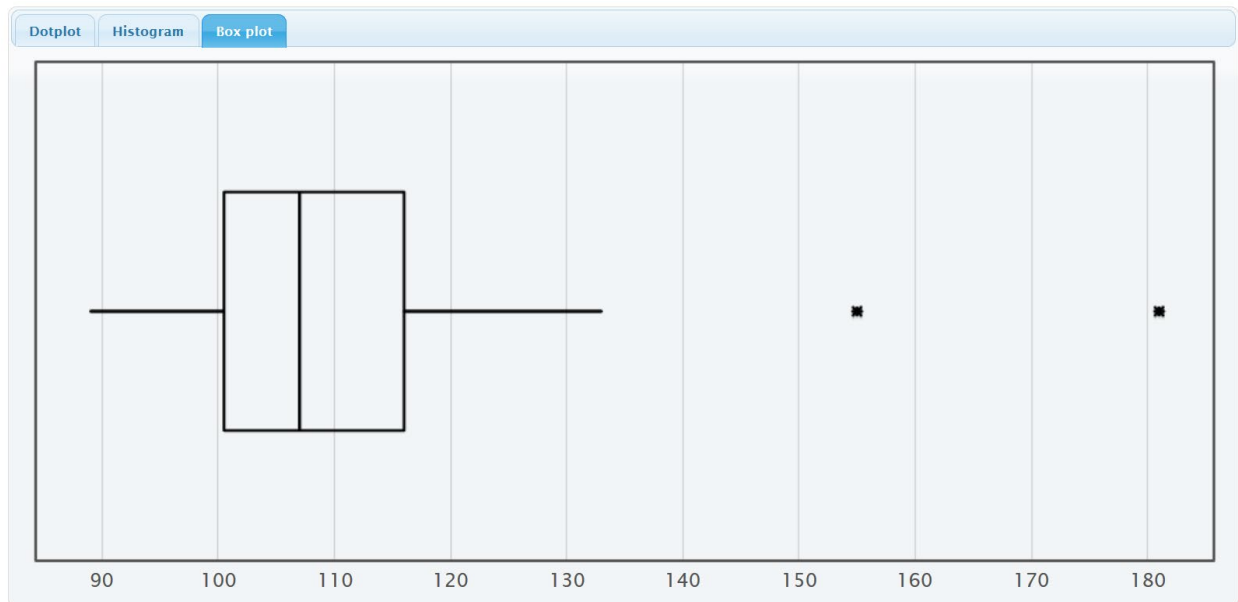
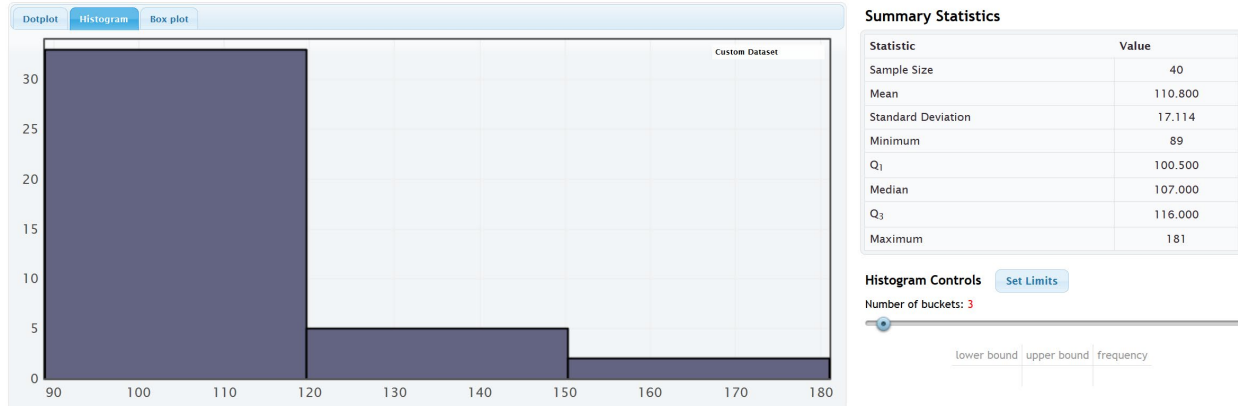
- The data is measuring the ages of bears. The units are months.
- There are 54 bears in the data set. (Sample size)
- The histogram shows that the data is skewed right.
- The youngest bear is 8 months old.
- The oldest bear is 177 months old.
- Since the data is not normal, we should use the median as our average or center. The median average is 34 months, so the average age of the bears is 34 months old.
- Since the data is not normal we should use the IQR as our best measure of typical spread. The IQR is the 3<sup>rd</sup> quartile minus the 1<sup>st</sup> quartile =  $58 - 17 = 41$  months. So typical bear ages are within 41 months of each other.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

- h) Since the data is not normal, typical values will fall between the 1<sup>st</sup> quartile (17 months) and the 3<sup>rd</sup> quartile (58 months). So typical bear are between 17 months old and 58 months old.
- i) The box plot has one star to the far right. This corresponds to the maximum value of 177 months. So there is only one high outlier in the data set at 177 months old.
- j) The box plot does not have any stars to the far left, so there is no low outliers in this data.

5.



- a) The data is measuring the systolic blood pressures of women. The units are millimeters of mercury (mm of Hg).
- b) There are 40 women in the data set. (Sample size)
- c) The histogram shows that the data is skewed right.
- d) The lowest systolic blood pressure for these women was 89 mm of Hg.
- e) The highest systolic blood pressure for these women was 181 mm of Hg.

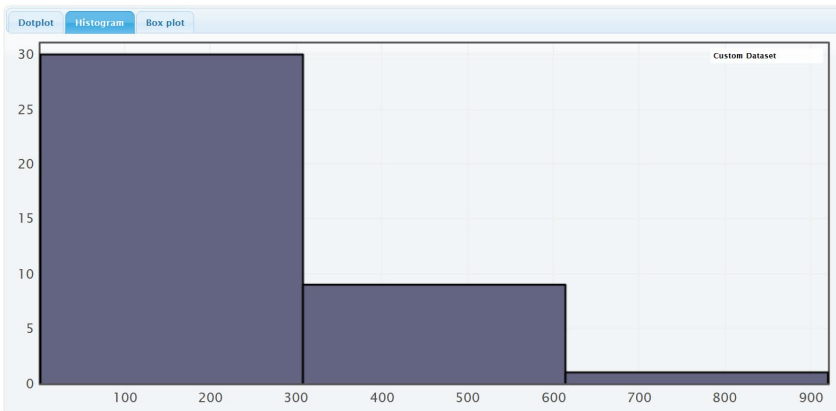
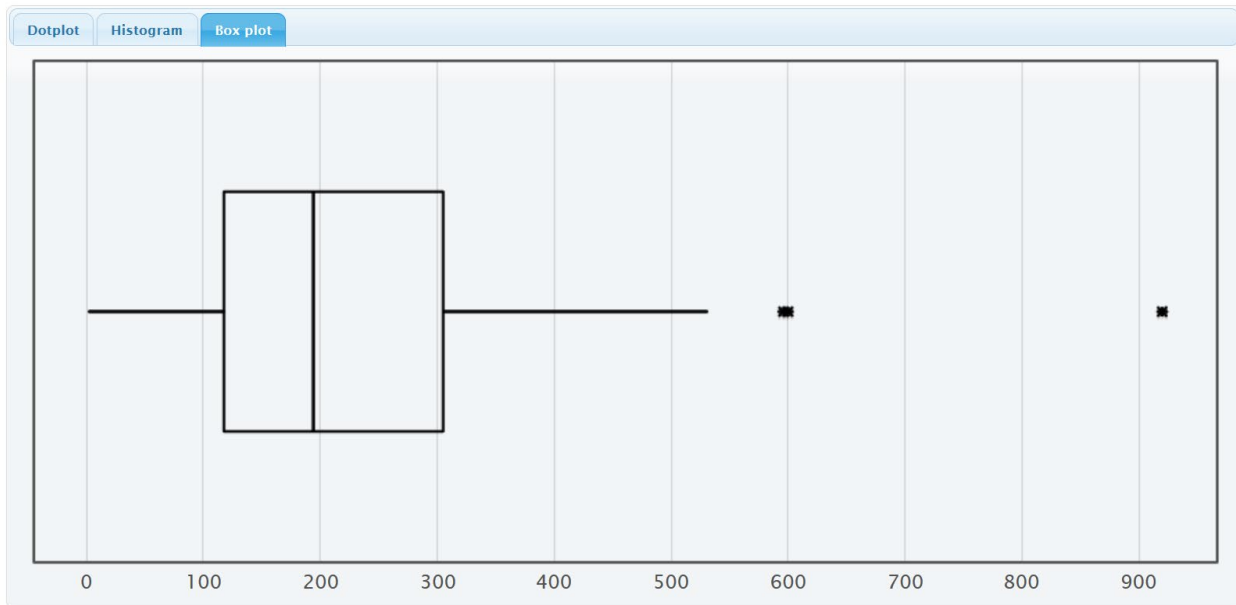




Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

- f) Since the data is not normal, we should use the median as our average or center. The median average is 107 mm of Hg, so the average systolic blood pressure for these women is 107 mm of Hg.
- g) Since the data is not normal we should use the IQR as our best measure of typical spread. The IQR is the 3<sup>rd</sup> quartile minus the 1<sup>st</sup> quartile =  $116 - 100.5 = 15.5$  mm of Hg. So typical women have a systolic blood pressure within 15.5 mm of Hg of each other.
- h) Since the data is not normal, typical values will fall between the 1<sup>st</sup> quartile (100.5 mm of Hg) and the 3<sup>rd</sup> quartile (116 mm of Hg). So typical women in this data had a systolic blood pressure between 100.5 mm of Hg and 116 mm of Hg.
- i) The box plot has two stars to the far right. This corresponds to 155 mm of Hg and the maximum value of 181 mm of Hg. So there is only two high outliers in the data set at 155 mm of Hg and 181 mm of Hg.
- j) The box plot does not have any stars to the far left, so there is no low outliers in this data.

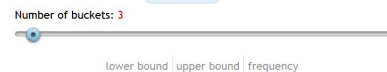
7.



**Summary Statistics**

Statistic	Value
Sample Size	40
Mean	240.875
Standard Deviation	185.982
Minimum	2
Q <sub>1</sub>	117.500
Median	194.000
Q <sub>3</sub>	305.000
Maximum	920

**Histogram Controls** [Set Limits](#)



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

- a) The data is measuring the cholesterol of women. The units are milligrams per deciliter (mg/dL).
- b) There are 40 women in the data set. (Sample size)
- c) The histogram shows that the data is skewed right.
- d) The lowest cholesterol for these women was 2 mg/dL. This may be a mistake in the data. Doesn't sound possible.
- e) The highest cholesterol for these women was 920 mg/dL.
- f) Since the data is not normal, we should use the median as our average or center. The median average is 194 mg/dL, so the average cholesterol for these women is 194 mg/dL.
- g) Since the data is not normal we should use the IQR as our best measure of typical spread. The IQR is the 3<sup>rd</sup> quartile minus the 1<sup>st</sup> quartile =  $305 - 117.5 = 187.5$  mg/dL. So typical women have a cholesterol within 187.5 mg/dL of each other.
- h) Since the data is not normal, typical values will fall between the 1<sup>st</sup> quartile (117.5 mg/dL) and the 3<sup>rd</sup> quartile (305 mg/dL). So typical women in this data had a cholesterol between 117.5 mg/dL and 305 mg/dL.
- i) The box plot has three stars to the far right. So there are three unusually high cholesterols for these women. They corresponds to the cholesterols of 596 mg/dL, 600 mg/dL and 920 mg/dL.
- j) The box plot does not have any stars to the far left, so there is no low outliers in this data.

9.

- a) The data is measuring the gas mileage for various cars. The units are in miles per gallon (mpg).
- b) There are 38 cars in the data set. ("N Total")
- c) The histogram shows that the data is skewed right.
- d) The lowest miles per gallon was 15.5 mpg.
- e) The highest miles per gallon was 37.3 mpg.
- f) Since the data is not normal, we should use the median as our average or center. The median average is 24.25 mpg, so the average gas mileage for these cars is 24.25 mpg.
- g) Since the data is not normal we should use the IQR as our best measure of typical spread. The IQR is given as 12.175 mpg. So typical cars in this data are within 12.175 mpg of each other.
- h) Since the data is not normal, typical values will fall between the 1<sup>st</sup> quartile (18.425 mpg) and the 3<sup>rd</sup> quartile (30.6 mpg). So typical cars in this data have a gas mileage between 18.425 mpg and 30.6 mpg.
- i) The box plot does not have any stars to the far right, so there is no high outliers in this data.
- j) The box plot does not have any stars to the far left, so there is no low outliers in this data.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

11.

- a) The data is measuring the horsepower for various cars. The units are the number of horsepower the car has.
- b) There are 38 cars in the data set. ("N Total")
- c) The histogram shows that the data is skewed right.
- d) The smallest horsepower for these cars was 65.
- e) The largest horsepower for these cars was 155.
- f) Since the data is not normal, we should use the median as our average or center. The median average is 100 horsepower, so the average for these cars is 100 horsepower.
- g) Since the data is not normal we should use the IQR as our best measure of typical spread. The IQR is given as 47.75 horsepower. So typical cars in this data are within 47.75 horsepower of each other.
- h) Since the data is not normal, typical values will fall between the 1<sup>st</sup> quartile (77.25 horsepower) and the 3<sup>rd</sup> quartile (125 horsepower). So typical cars in this data have a horsepower between 77.25 and 125.
- j) The box plot does not have any stars to the far right, so there is no high outliers in this data.
- j) The box plot does not have any stars to the far left, so there is no low outliers in this data.

13.

- a) Q1 is a measure of position.
  - b) Mean is a measure of center.
  - c) Variance is a measure of spread.
  - d) Standard deviation is a measure of spread.
  - e) Minimum value is a measure of position.
  - f) Q3 is a measure of position.
  - g) The mode is a measure of center.
  - h) The IQR is a measure of spread.
  - i) The median is a measure of center.
  - j) The range is a measure of spread.
  - k) The maximum is a measure of position.
  - l. The midrange is a measure of center.
- 



## 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 5<sup>th</sup> 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.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

**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.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

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)  $11/74 \approx 0.149$
- b)  $0.149 \times 100\% = 14.9\%$

Approximately 14.9% of the company are managers.

- c)  $27/74 \approx 0.365$
- d)  $0.365 \times 100\% = 36.5\%$

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

- e)  $36/74 \approx 0.486$
- f)  $0.486 \times 100\% = 48.6\%$

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

g) Percent of Increase =  $(0.365 - 0.149) / 0.149 \approx 145.0\%$  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.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

h) Percent of Increase =  $(0.486 - 0.365) / 0.365 \approx 33.2\%$  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.

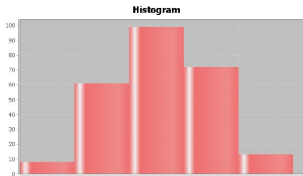
11.

$60\% = 0.6$

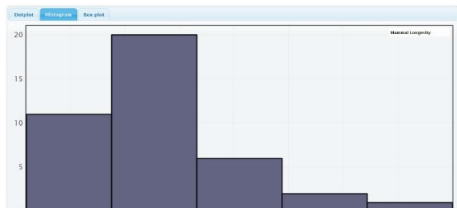
Estimated Amount =  $0.6 \times 41743 \approx 25,046$  voters in Saugus.

12.

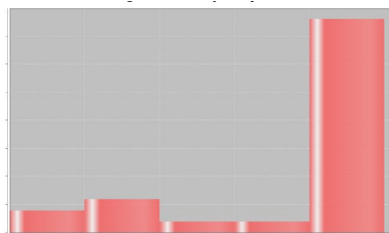
a) A normal or normally distributed histogram is unimodal and symmetric. This means that we expect the highest bar or bars to be in the middle with smaller and smaller bars as we go away from the middle. The left and right tails will be approximately the same length.



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.



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

- 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 ( $Q_3$ ) 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<sup>th</sup> percentile or 2<sup>nd</sup> quartile ( $Q_2$ ) 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  $Q_3$ . ( $Q_3 + (1.5 \times \text{IQR})$ ).
- 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  $Q_1$ . ( $Q_1 - (1.5 \times \text{IQR})$ ).
- 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.





Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

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) =  $7.345 + (2 \times 1.376) = 10.097$ . So any employees that have been with the company 10.097 years or more is 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) =  $7.345 - (2 \times 1.376) = 4.593$ . 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



Introduction to Statistics for Community College Students  
Appendix A: Answer Keys to Odd Exercises Chapter 1

e)  $0.018 \times 100\% = 1.8\%$

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.

---



*This chapter is from Introduction to Statistics for Community College Students,  
1<sup>st</sup> Edition, by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed  
under a "CC-By" [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) – 10/1/18*