AP Statistics Summer Assignment 2025

Due the First Day of Class!!!!

Welcome to AP Statistics!

The purpose of this summer assignment is to:

1. Refresh your knowledge on statistics.
2. Demonstrate your abilities to analyze data and determine conclusions.

All of your work is attached with instructions. Be sure to finish all work by the beginning of school in August as it will be collected and graded. There will also be an assessment on the materials in this packet within the first two weeks of returning to school.

If you are struggling and need more help with any topics, there is a free online Statistical tutoring site that will give you additional information, www.stattrek.com

Upon your return to school in August, this packet is expected to be turned in the first day of class. You are expected to complete each part of the problems and to construct all data displays neatly. These assignments will count as two grades as follows:

1. **Homework Grade: Complete ALL sections of the notes packet – 30 pts**
2. **Quiz Grade: Complete ALL of the worksheets showing all work/explanations as needed. We will be grading it together in class. – 70 pts**

<u>Summer Outline:</u>

1. **Lesson 1:**
    a) Watch the following video and complete the What is Statistics Notes in the notes packet: https://www.youtube.com/watch?v=PPEnQjp7tak&t=20s

    b) You will have a Canvas quiz over this the first day of class, you will be able to use your completed notes.

2. **Lesson 2:**
    a) Watch the following video and complete the 1.0 Introduction Notes in the notes packet:
    https://youtu.be/3oOKsI3VpuI

    b) Complete the 1.0 Introduction Worksheet in the Worksheet Packet.

3. **Lesson 3:**
   a) Watch the following video and complete the 1.1 & 1.2 Displaying Data with Graphs Notes in the notes packet:
      https://youtu.be/JavLiovLZT4

   b) Complete the 1.1 & 1.2 Displaying Data with Graphs Worksheet in the Worksheet Packet.

   c) Watch the following video and complete Example: Create a frequency table and histogram for the set of data in the notes packet:
      https://youtu.be/vLunrTIn1fM

   d) Complete the Histogram Worksheet in the Worksheet Packet.

   e) Watch the following video and complete the 1.2 Part 2: Describing Distributions Notes in the notes packet:
      https://youtu.be/11FhMcwqD1k

   f) Complete the 1.2 Part 2 Describing Distributions Worksheet in the Worksheet Packet.

4. **Lesson 4:**
   a) Watch the following videos and complete the 1.3 Describing Quantitative Data with Numbers Notes in the notes packet:
      https://youtu.be/sMiAEUEG214
      https://www.youtube.com/watch?v=O2vfa3mjMKI

   b) Complete the 1.3 Describing Distributions Worksheet in the Worksheet Packet.

AP Statistics

Chapter 1

Lecture Notes

What is Statistics?

Statistics is the science of …

- Statistics is the science of _____ from data.

- It is concerned with the _____,

  _____, and _____ of data,

  as well as the effective _____ and

  _____ of results relying on data.

- This branch of mathematics allows us to make _____ reasoning necessary for important advances in the sciences, such as medicine and genetics, and for making important _____ in business and public policy.

- Basic definition of statistics: Statistics is the science of:

  _____

  _____

  _____

- The two branches of statistics are:

- Descriptive Statistics

- Inferential Statistics

- Population

- Sample

- Generally, it is _____, if not impossible, to collect, analyze and draw conclusions from a _____. Therefore, we use statistics to study _____ that are representative of the population to make _____ and _____ about the population itself.

- The study of statistics deals with understanding how to obtain _____ and work with the sample data to make statistically justified _____.

Section 1.0: Introduction

Instructional Objectives:	Students will be able to:
	(1) Identify the individuals and variables in a set of data.
	(2) Classify variables as categorical or quantitative.

Organizing Data

Vocabulary:

- **Individual:** is an _____ described in a set of data. Individuals can be _____, _____, or _____.

- **Variable:** an attribute that can take different _____ for _____ individuals.

  o **Categorical Variable:** assigns _____ that place each individual into a particular _____, called a _____.

  o **Quantitative Variable:** takes _____ values that are quantities—_____ or _____.

    - **Discrete Variable:** A quantitative variable that takes a _____ set of possible values with _____ between them.

    - **Continuous Variable:** A quantitative variable that can take ____ _____ in an interval on the number line.

Analyzing Data

Vocabulary:

- **Distribution:** The distribution of a variable tells us what _____ the variable takes and _____ often it takes those values.

How to Analyze Data:

1. _____ each variable by itself.
2. Then _____ relationships among the variables.
   - Start with _____
   - Add _____ summaries

From Data Analysis to Inference:

_____ **data** from a representative

_____ _statistics_

_____...

_____ _statistics_

Perform _____ **analysis...**

Make an _____ about the _____.

_____ _statistics_
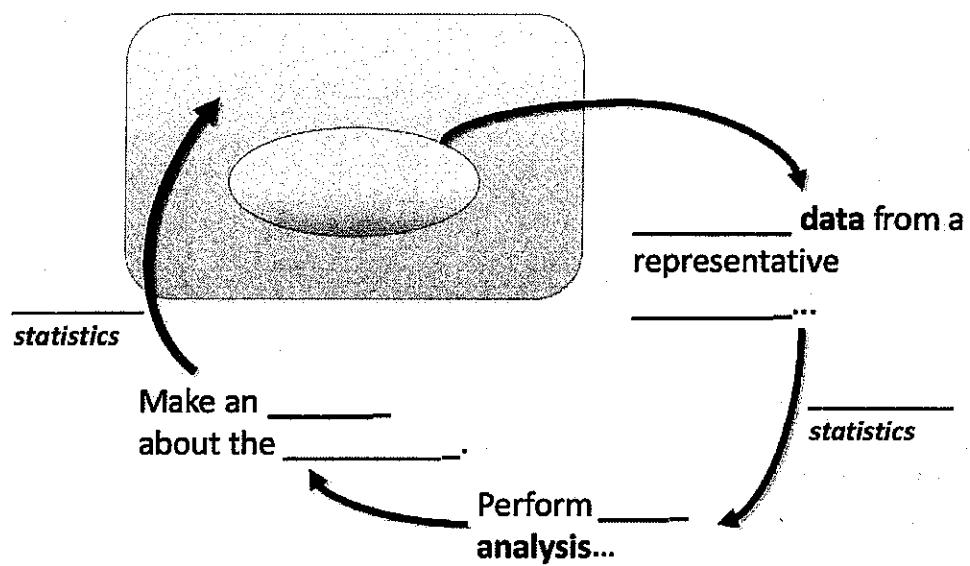
1.1 & 1.2: Displaying Data with Graphs

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Graphs for Categorical Data:

1. _____ Graph:

Frequency table	
Preferred method	**Frequency**
Cell phone	2
Facebook	1
In person	3
Text messaging	4

Preferred communication method
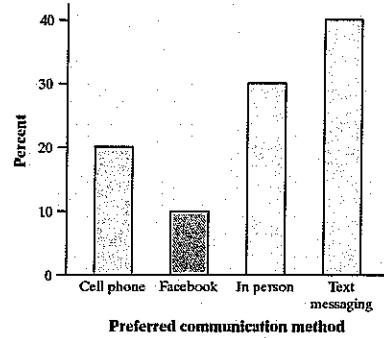
2. _____ Chart:

| Relative frequency table	
Preferred method	**Relative frequency**
Cell phone	2/10 = 0.20 or 20%
Facebook	1/10 = 0.10 or 10%
In person	3/10 = 0.30 or 30%
Text messaging	4/10 = 0.40 or 40%
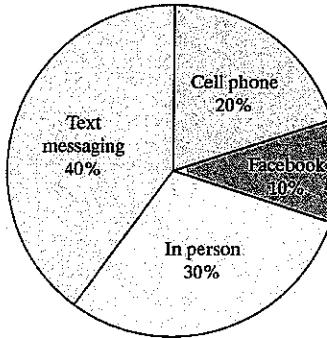
Preferred communication method

3. A **side-by-side** _____ **graph** displays the distribution of a categorical variable for each value of another categorical variable. The bars are grouped together based on the values of one of the categorical variables and placed side by side.
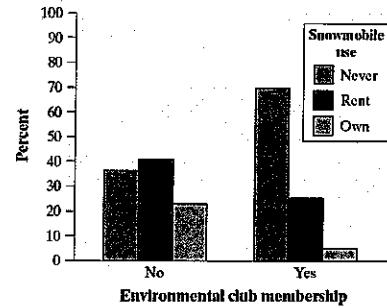
4. A _____ **bar graph** displays the distribution of a categorical variable as segments of a rectangle, with the area of each segment proportional to the percent of individuals in the corresponding category.
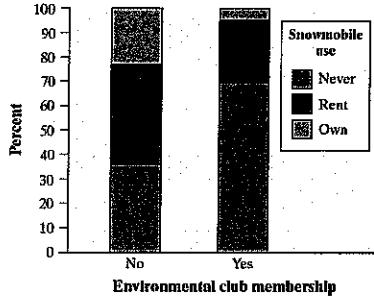
5. A _____ **plot** is a modified segmented bar graph in which the width of each rectangle is proportional to the number of individuals in the corresponding category.
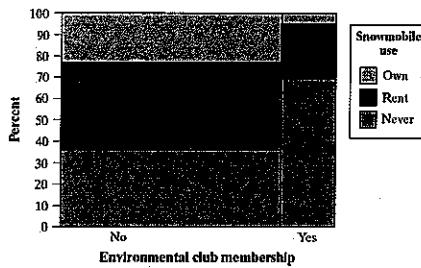
Relationships Between Two Categorical Variables

> There is an _____ between two variables if _____ the value of one variable helps us predict the value of the other.

> If knowing the value of one variable _____ _____ help us predict the value of the other, then there is _____ **association** between the variables.

Graphs for Quantitative Data:

1. A _____ shows each data value as a dot above its location on a number line.

   ✓ Draw a horizontal axis (a number line) and _____ it with the variable name.
   ✓ _____ the axis from the minimum to the maximum value.
   ✓ Mark a dot above the location on the horizontal axis corresponding to each data value.

2. A _____ shows each data value separated into two parts: a stem, which consists of all but the final digit, and a leaf, the final digit. The stems are ordered from lowest to highest and arranged in a vertical column. The leaves are arranged in increasing order out from the appropriate stems.

   ✓ Separate each observation into a stem, consisting of all but the final digit, and a leaf, the final digit. Write the stems in a vertical column with the smallest at the top. Draw a vertical line at the right of this column.
   ✓ Write each leaf in the row to the right of its stem.
   ✓ Arrange the leaves in increasing order out from the stem.
   ✓ Provide a key that identifies the variable and explains what the stems and leaves represent.

| 50 | 26 | 26 | 31 | 57 | 19 | 24 | 22 | 23 | 38 |
| 13 | 50 | 13 | 34 | 23 | 30 | 49 | 13 | 15 | 51 |

When data values are "bunched up", we can get a better picture of the distribution by **splitting stems**.

Two distributions of the same quantitative variable can be compared using a **back-to-back stemplot** with common stems.

| 50 | 26 | 26 | 31 | 57 | 19 | 24 | 22 | 23 | 38 |
| 13 | 50 | 13 | 34 | 23 | 30 | 49 | 13 | 15 | 51 |

| 14 | 7 | 6 | 5 | 12 | 36 | 8 | 7 | 10 | 10 |
| 10 | 11 | 4 | 5 | 22 | 7 | 5 | 10 | 35 | 7 |

3. A _____ shows each interval of values as a bar. The heights of the bars show the frequencies or relative frequencies of values in each interval.

How to make a histogram:

1) Choose _____-width intervals that span the data.
2) Make a table that shows the frequency or relative frequency of individuals in each interval.
3) Draw horizontal and vertical axes. _____ the axes.
4) _____ the axes.
5) Draw bars above the intervals. The bar heights correspond to the frequency or relative frequency of individuals in that interval.
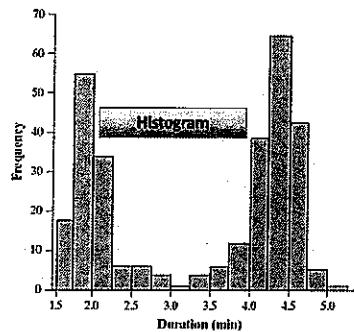
State	Percent	State	Percent	State	Percent
Alabama	2.8	Louisiana	2.9	Ohio	3.6
Alaska	7.0	Maine	3.2	Oklahoma	4.9
Arizona	15.1	Maryland	12.2	Oregon	9.7
Arkansas	3.8	Massachusetts	14.1	Pennsylvania	5.1
California	27.2	Michigan	5.9	Rhode Island	12.6
Colorado	10.3	Minnesota	6.6	South Carolina	4.1
Connecticut	12.9	Mississippi	1.8	South Dakota	2.2
Delaware	8.1	Missouri	3.3	Tennessee	3.9
Florida	18.9	Montana	1.9	Texas	15.9
Georgia	9.2	Nebraska	5.6	Utah	8.3
Hawaii	16.3	Nevada	19.1	Vermont	3.9
Idaho	5.6	New Hampshire	5.4	Virginia	10.1
Illinois	13.8	New Jersey	20.1	Washington	12.4
Indiana	4.2	New Mexico	10.1	West Virginia	1.2
Iowa	3.8	New York	21.6	Wisconsin	4.4
Kansas	6.3	North Carolina	6.9	Wyoming	2.7
Kentucky	2.7	North Dakota	2.1		

Frequency Table

Class

0 to <5

5 to <10

10 to <15

15 to <20

20 to <25

25 to <30

Total

Example: Create a frequency table and histogram for the set of data.

Jeff is the branch manager at a local bank. Recently, Jeff's been receiving customer feedback saying that the wait times for a client to be served by a customer service representative are too long. Jeff decides to observe and write down the time spent by each customer on waiting. Here are his findings from observing and writing down the wait times, in seconds, spent by 25 customers:

43, 36, 38, 37, 45, 40, 50, 62, 47, 31, 72, 42, 46, 30, 31, 36, 45, 54, 46, 37, 43, 60, 25, 33

Frequency table:

Interval	# of values
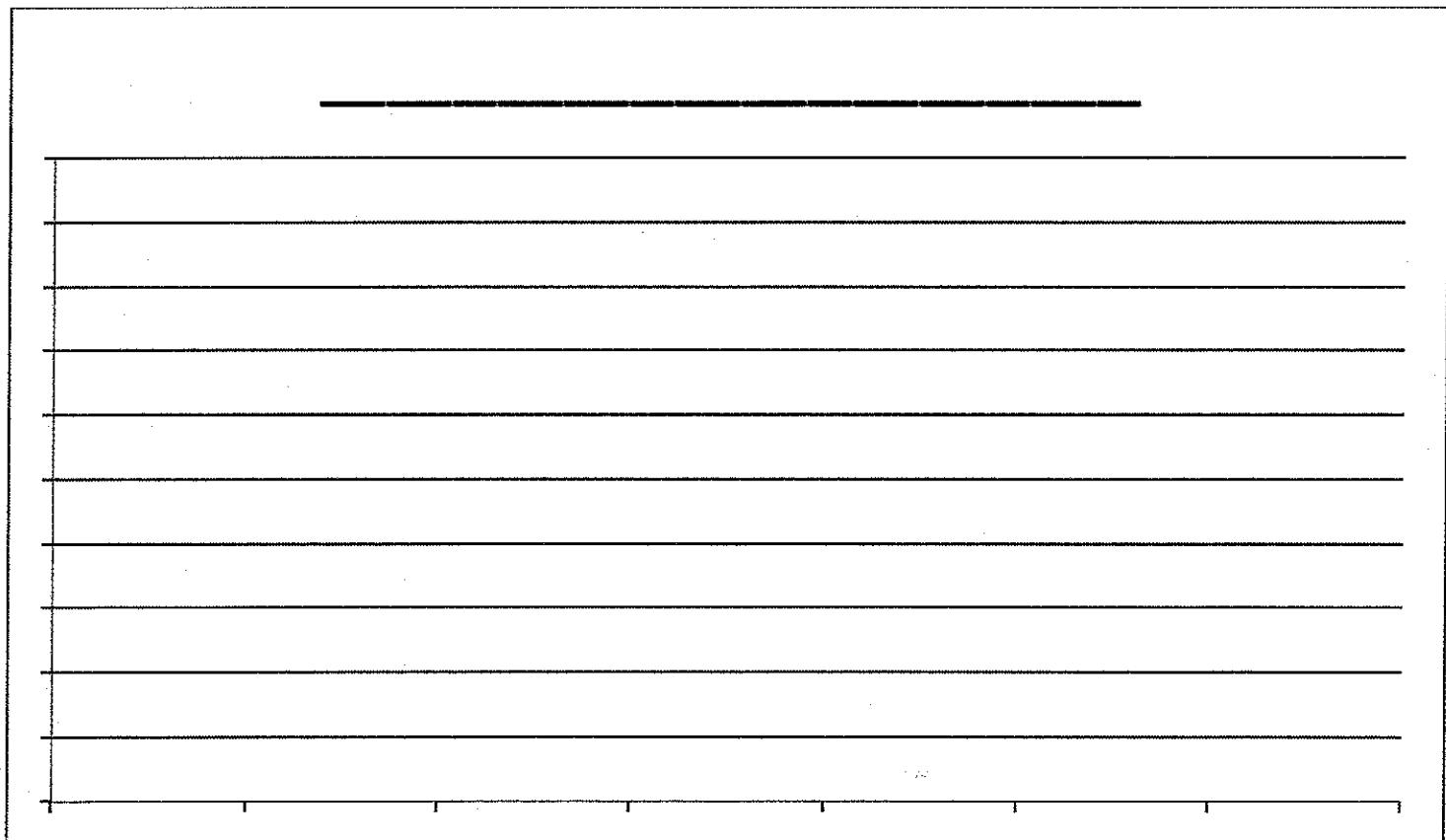
Section 1.2 part 2: Describing Distributions

Instructional Objectives: Students will be able to:
(1) Identify the shape of a distribution by its graph.
(2) Describe the overall pattern (shape, center and variability) of a distribution and Identify any major departures from the pattern (outliers).
(3) Compare distributions of quantitative data using dotplots, stemplots, and histograms.

Describing Shape

✓ A distribution is roughly _____ if the right side of the graph (containing the half of observations with the largest values) is approximately a mirror image of the left side.

✓ A distribution is _____ **to the** _____ if the right side of the graph is much longer than the left side.

✓ A distribution is _____ **to the** _____ if the left side of the graph is much longer than the right side.

✓ The distribution of a quantitative variable is _____ if it has a single peak.

✓ The distribution of a quantitative variable is _____ if it has two distinct clusters and peaks.

✓ The distribution of a quantitative variable is _____ _____ if the frequencies are about the same for all values.

How to Describe the Distribution of a Quantitative Variable

In any graph, look for the _____ _____ and for clear _____ from that pattern.

An important kind of departure is an _____, an observation that falls outside the overall pattern.

Describe the distribution of goals scored in 20 games played by the 2016 U.S. women's soccer team.

Shape:

Outliers:

Goals scored
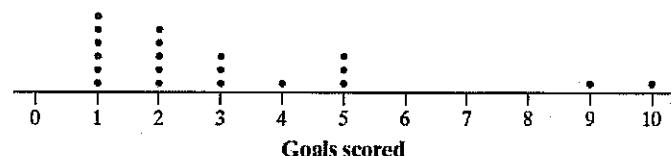
Center:

Variability:

Comparing Distributions

We used Census At School's "Random Data Selector" to choose 50 students from each country. Here are dotplots of the household sizes reported by the survey respondents. Compare the distributions of household size for these two countries.

Shape:
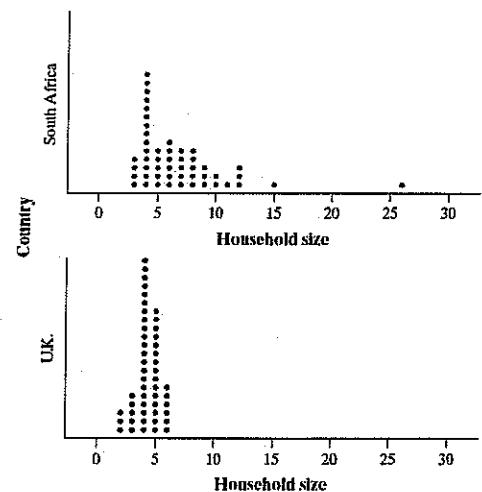
Outliers:

Center:

Variability:

Section 1.3: Describing Quantitative Data with Numbers

Instructional Objectives: Students will be able to:

(1) Calculate measures of center (mean, median) for a distribution of quantitative data.

(2) Calculate and Interpret measures of variability (range, standard deviation, IQR) for a distribution of quantitative data.

(3) Explain how outliers and skewness affect measure of center and variability.

(4) Identify outliers using the 1.5 x IQR Rule

(5) Make and Interpret boxplots of quantitative data.

(6) Use boxplots and numerical summaries to compater distributions of quantitative data.

Measuring Center

Mean:

✓ The _____ of a distribution of quantitative data is the _____ of all the individual data values.

✓ To find the mean, _____ all the values and _____ by the _____ number of observations.

✓ If the ___ observations are $x_1, x_2, ..., x_n$, the sample mean ___ (pronounced " x-bar ") is given by the following formula:

✓ The symbol ___ refers to the mean of a _____.

✓ The notation ___ refers to the mean of a _____.

✓ A _____ is a number that describes some characteristic of a _____.

✓ A _____ is a number that describes some characteristic of a _____.

Example: Here are the data on the number of goals scored in 20 games played by the 2016 U.S. women's soccer team:

$$5\ 5\ 1\ 10\ 5\ 2\ 1\ 1\ 2\ 3\ 3\ 2\ 1\ 4\ 2\ 1\ 2\ 1\ 9\ 3$$

✓ A statistical measure is _____ if it isn't sensitive to extreme values.

✓ The mean is sensitive to extreme values in a distribution. The mean is _____ a resistant measure of center.

Median:

✓ The _____ is the _____ of a distribution, the number such that about _____ the observations are smaller and about half are larger.

✓ To find the median, arrange the data values from smallest to largest.

- If the number ___ of data values is ___, the median is the middle value in the ordered list.

- If the number ___ of data values is _____, the median is the average of the two middle values in the ordered list.

Example: Here are the highway fuel economy ratings for a sample of 25 model year 2018 Toyota 4Runners tested by the EPA:

22.4, 22.4, 22.3, 23.3, 22.3, 22.3, 22.5, 22.4, 22.1, 21.5, 22.0, 22.2, 22.7,

22.8, 22.4, 22.6, 22.9, 22.5, 22.1, 22.4, 22.2, 22.9, 22.6, 21.9, 22.4

Example: Here are the data on the number of goals scored in 20 games played by the 2016 U.S. women's soccer team:

5 5 1 10 5 2 1 1 2 3 3 2 1 4 2 1 2 1 9 3

Comparing the Mean and Median:

✓ Skewed to the Left:

✓ Skewed to the Right:

✓ Roughly Symmetric:
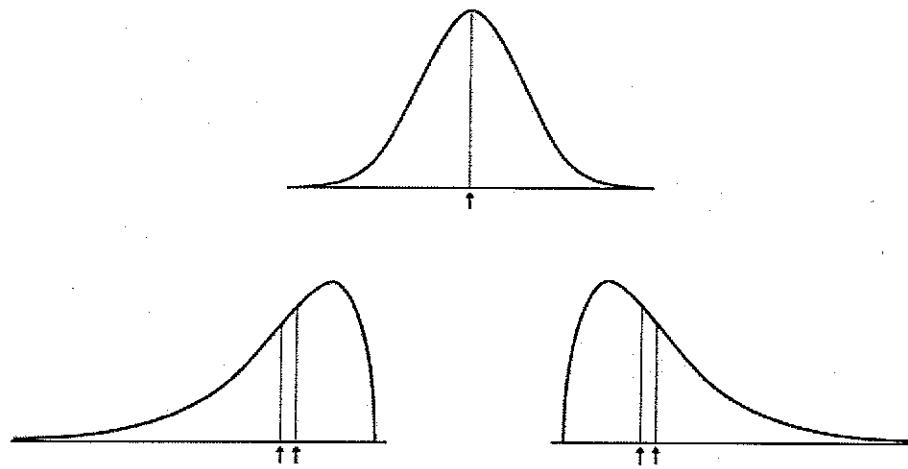
Effect of Skewness and Outliers on Measures of Center:

✓ If a distribution of quantitative data is roughly _____ and has ___ outliers, the mean and median will be _____.

✓ If the distribution is strongly _____, the _____ will be pulled in the direction of the _____ but the _____ won't.

   • For a _____-skewed distribution, we expect the mean to be _____ than the median.

   • For a ____-skewed distribution, we expect the mean to be _____ than the median.

✓ The _____ is _____ to outliers but the mean isn't.

Measuring Variability

Range:

✓ The _____ of a distribution is the distance between the _____ value and the _____ value. That is,

Example: Here are the data on the number of goals scored in 20 games played by the 2016 U.S. women's soccer team:

$$5\ 5\ 1\ 10\ 5\ 2\ 1\ 1\ 2\ 3\ 3\ 2\ 1\ 4\ 2\ 1\ 2\ 1\ 9\ 3$$

- The range of a data set is a _____ number.

- The range is _____ a resistant measure of variability.

Standard Deviation:

✓ The **standard deviation** measures the typical _____ of the values in a distribution from the

_____.

✓ **How to calculate standard deviation and variance:**
  1. Find the mean of the distribution.
  2. Calculate the *deviation (*value – mean) of each value from the mean.
  3. Square each of the deviations.
  4. Add all the squared deviations, divide by $n - 1$. This is the **sample variance**.
  5. Take the square root. This is the **sample standard deviation**.

✓ ____ = sample standard deviation

$$s_x = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

✓ ____ **This value is known as the sample variance.**

Example: Eleven high school students were asked how many "close" friends they have. Here are their responses: $1\ 2\ 2\ 2\ 3\ 3\ 3\ 3\ 4\ 4\ 6$

Properties of Standard Deviation:

- s_x is always greater than or equal to _____.

- _____ values of s_x indicate greater variation.

- s_x is ___ a _____ measure of variability.

- s_x measures _____ about the _____.

The Interquartile Range (IQR):

✓ The **quartiles** of a distribution divide the ordered data set into _____ groups having roughly the _____ number of values.

✓ To find the quartiles, arrange the data values from _____ to _____ and find the _____.

✓ The **first quartile** ____ is the _____ of the data values that are to the _____ of the median in the ordered list.

✓ The **third quartile** ____ is the median of the data values that are to the _____ of the _____ in the ordered list.

✓ The **interquartile range** ____ is the distance between the ____ and _____ quartiles of a distribution. In symbols:

Amount collected ($)
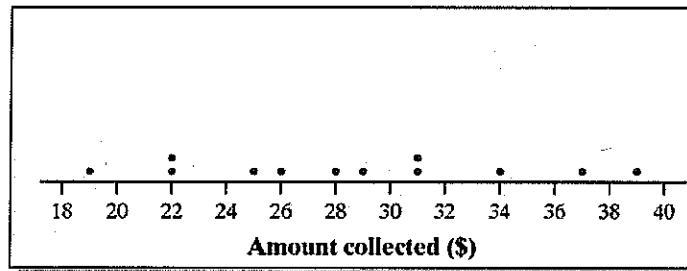
Example:

Travel times to work for 20 randomly selected New Yorkers

10	30	5	25	40	20	10	15	30	20	15	20	85	15	65	15	60	60	40	45

Identifying Outliers

✓ Although there are several rules for outliers, one of the most common rules is the

✓ How to identify outliers:

Example:

Highway fuel economy ratings for twenty-five 2018 Toyota 4Runners tested by the EPA:

21.5 21.9 22.0 22.1 22.1 22.2 22.2 22.3 22.3 22.3 22.4 22.4 22.4

22.4 22.4 22.4 22.5 22.5 22.6 22.6 22.7 22.8 22.9 22.9 23.3

Q_1 = 22.2 mpg

Q_3 = 22.6 mpg

IQR = 0.4 mpg

Highway fuel economy (mpg)
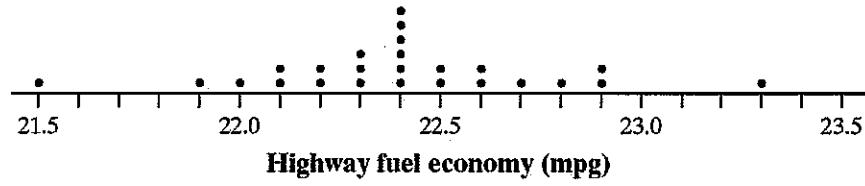
Identifying Outliers

Why look for outliers?

1. They might be _____ data values.

2. They can indicate a remarkable occurrence.

3. They can heavily _____ the values of some summary statistics, like the mean, range, and standard deviation.

Making and Interpreting Boxplots

✓ The ____-**number summary** of a distribution of quantitative data consists of the _____, the _____ quartile Q_1, the _____, the _____ quartile Q_3, and the _____.

Highway fuel economy (mpg)
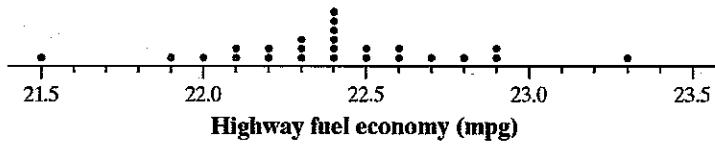
✓ A _____ is a visual representation of the five-number summary.

✓ How to make a Boxplot:
1. Find the _____-number summary.
2. Identify _____ using the 1.5 × IQR rule.
3. Draw and label the horizontal _____.
4. _____ the axis.
5. Draw a ____.
6. Mark the _____.
7. Draw _____.

Highway fuel economy (mpg)
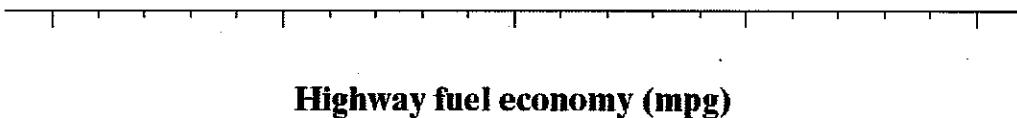
AP Statistics

Chapter 1

Worksheets

Name: _____

Answer the following questions completely.

1. Here is a small part of the data set that describes the students in an AP®® Statistics class. The data come from anonymous responses to a questionnaire filled out on the first day of class.

Gender	Grade level	GPA	Children in family	Homework last night (min)	Android or iPhone?
F	9	2.3	3	0-14	iPhone
M	11	3.8	6	15-29	Android
M	10	3.1	2	15-29	Android
F	10	4.0	1	45-59	iPhone
F	10	3.4	4	0-14	iPhone
F	10	3.0	3	45-59	Android
M	9	3.9	2	15-29	iPhone
M	12	3.5	2	0-14	iPhone

a. Identify the individuals in this data set.

b. What are the variables? Classify each as categorical or quantitative.

2. Many people like to ride roller coasters. Amusement parks try to increase attendance by building exciting new coasters. The following table displays data on several roller coasters that were opened in a recent year.

Roller coaster	Type	Height (ft)	Design	Speed (mph)	Duration (sec)
Wildfire	Wood	187.0	Sit down	70.2	120
Skyline	Steel	131.3	Inverted	50.0	90
Goliath	Wood	165.0	Sit down	72.0	105
Helix	Steel	134.5	Sit down	62.1	130
Banshee	Steel	167.0	Inverted	68.0	160
Black Hole	Steel	22.7	Sit down	25.5	75

a. Identify the individuals in this data set.

b. What are the variables? Classify each as categorical or quantitative.

3. What measures can be taken, especially when restoring historic wooden buildings, to help wood surfaces resist weathering? In a study of this question, researchers prepared wooden panels and then exposed them to the weather. Some of the variables recorded were type of wood (yellow poplar, pine, cedar); type of water repellent (solvent-based, water-based); paint thickness (millimeters); paint color (white, gray, light blue); weathering time (months).

Which of the following is a true statement?

a) Weathering time is a quantitative variable.
b) Type of water repellent is a quantitative variable.
c) Paint thickness is a categorical variable.
d) Type of wood is a quantitative variable.
e) Paint color is a quantitative variable.

4. At the Census Bureau website, you can view detailed data collected by the American Community Survey. The following table includes data for 10 people chosen at random from the more than 1 million people in households contacted by the survey. "School" gives the highest level of education completed.

Weight (lb)	Age (years)	Travel to work (min)	School	Gender	Income last year ($)
187	66	0	Ninth grade	1	24,000
158	66	n/a	High School grad	2	0
176	54	10	Assoc. degree	2	11,900
339	37	10	Assoc. degree	1	6,000
91	27	10	Some college	2	30,000
155	18	n/a	High school grad	2	0
213	38	15	Master's degree	2	125,000
194	40	0	High school grad	1	800
221	18	20	High school grad	1	2,500
193	11	n/a	Fifth grade	1	0

a. The individuals in this data set are:

   a) 10 people
   b) households
   c) 1 million people
   d) age, travel to work, school, gender, income
   e) the census bureau

b. This data set contains

   a) 6 variables, 2 of which are categorical
   b) 7 variables, 2 of which are categorical
   c) 7 variables, 1 of which is categorical
   d) 6 variables, 1 of which is categorical
   e) None of these.

5. A college admissions office collects data from each incoming freshman on several quantitative variables: distance from their home to campus, number of siblings, how many books they have read in the past month, and how long it took them to complete an online survey. Classify each variable as discrete or continuous.

6. A social media company records data from each of its users on several quantitative variables: time spent on the site, how many times they visited the site, number of likes received, and how long since they created a member profile. Classify each variable as discrete or continuous.

1. ACCIDENTAL DEATHS

In 1997 there were 92,353 deaths from accidents in the United States. Among these were 42,340 deaths from motor vehicle accidents, 11,858 from falls, 10,163 from poisoning, 4051 from drowning, and 3601 from fires. The rest were listed as "other" causes.

a. Find the percent of accidental deaths from each of these causes, rounded to the nearest percent.

b. What percent of accidental deaths were from "other" causes?

c. NEATLY create a well-labeled bar graph of the distribution of causes of accidental deaths. Be sure to include an "other causes" bar and provide labels.
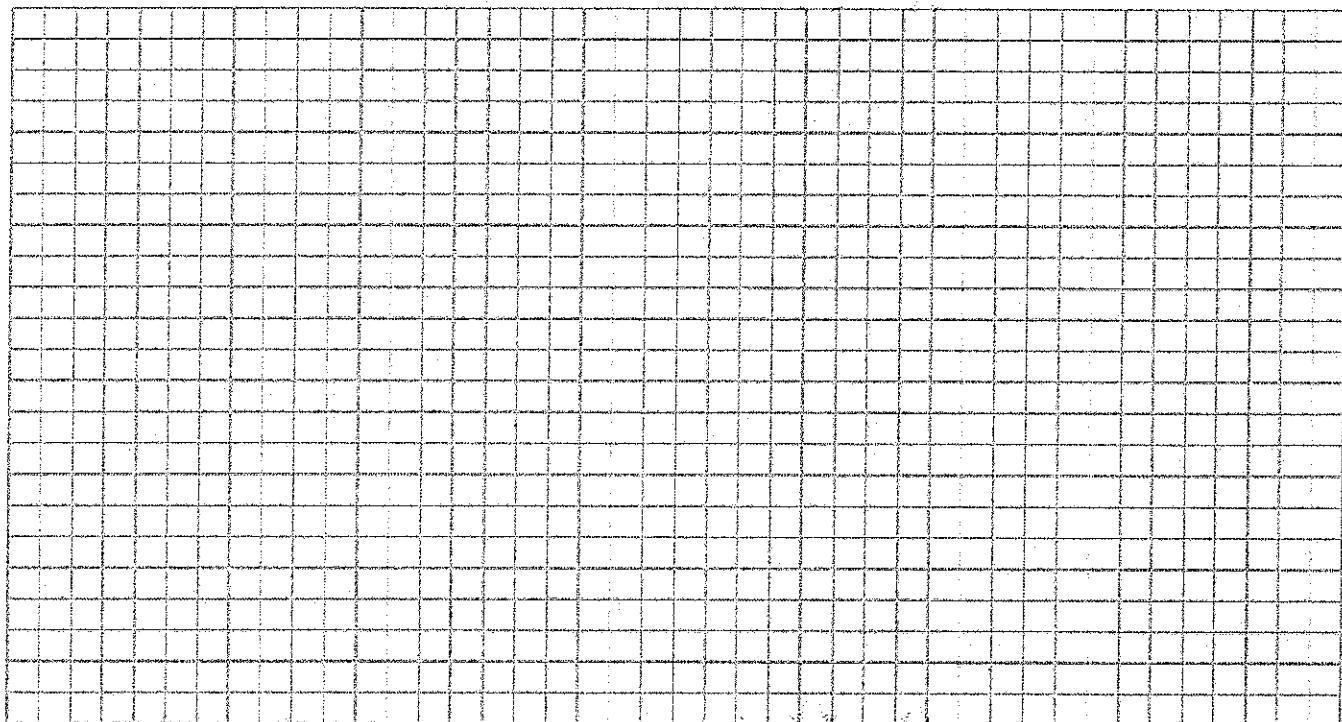
d. A pie chart is another graphical display used to show all the categories in a categorical variable relative to each other. Create a pie chart for the accidental death percentages.

2. IT'S A TWISTA

The data below gives the number of hurricanes that happened each year from 1944 through 2000 as reported by Science magazine.

3	2	1	4	3	7	2	3	3	2	5	2	2	4	2	2	6	0	2	5	1	3	1	0
3	2	1	0	1	2	3	2	1	2	2	2	3	1	1	1	3	0	1	3	2	1	2	1
1	0	5	6	1	3	5	3																

Make a dotplot to display these data. Make sure you include appropriate labels, title, and scale.

3. SHOPPING SPREE!

A marketing consultant observed 50 consecutive shoppers at a supermarket. One variable of interest was how much each shopper spent in the store. Here are the data (round to the nearest dollar), arranged in increasing order:

3	9	9	11	13	14	15	16	17	17
18	18	19	20	20	20	21	22	23	24
25	25	26	26	28	28	28	28	32	35
36	39	39	41	43	44	45	45	47	49
50	53	55	59	61	70	83	86	86	93

Make a stemplot using tens of dollars as the stem and dollars as the leaves. Make sure you include appropriate labels, title and key.

4. WHERE DO OLDER FOLKS LIVE?

This table gives the percentage of residents aged 65 or older in each of the 50 states.

| State	Percent	State	Percent	State	Percent
Alabama	13.1	Louisiana	11.5	Ohio	13.4
Alaska	5.5	Maine	14.1	Oklahoma	13.4
Arizona	13.2	Maryland	11.5	Oregon	13.2
Arkansas	14.3	Massachusetts	14.0	Pennsylvania	15.9
California	11.1	Michigan	12.5	Rhode Island	15.6
Colorado	10.1	Minnesota	12.3	South Carolina	12.2
Connecticut	14.3	Mississippi	12.2	South Dakota	14.3
Delaware	13.0	Missouri	13.7	Tennessee	12.5
Florida	18.3	Montana	13.3	Texas	10.1
Georgia	9.9	Nebraska	13.8	Utah	8.8
Hawaii	13.3	Nevada	11.5	Vermont	12.3
Idaho	11.3	New Hampshire	12.0	Virginia	11.3
Illinois	12.4	New Jersey	13.6	Washington	11.5
Indiana	12.5	New Mexico	11.4	West Virginia	15.2
Iowa	15.1	New York	13.3	Wisconsin	13.2
Kansas	13.5	North Carolina	12.5	Wyoming	11.5
Kentucky	12.5	North Dakota	14.4		

Histograms are a way to display groups of quantitative data into bins (the bars). These bins have the same width and scale and are touching because the number line is continuous. To make a histogram you must first decide on an appropriate bin width and count how many observations are in each bin. The bins for percentage of residents aged 65 or older have been started below for you.

5. Finish the chart of Bin widths and then create a histogram using those bins on the grid below. Make sure you include appropriate labels, title and scale.

Bin Widths	Frequency
4 to < 6	1
6 to < 8	
8 to < 10	
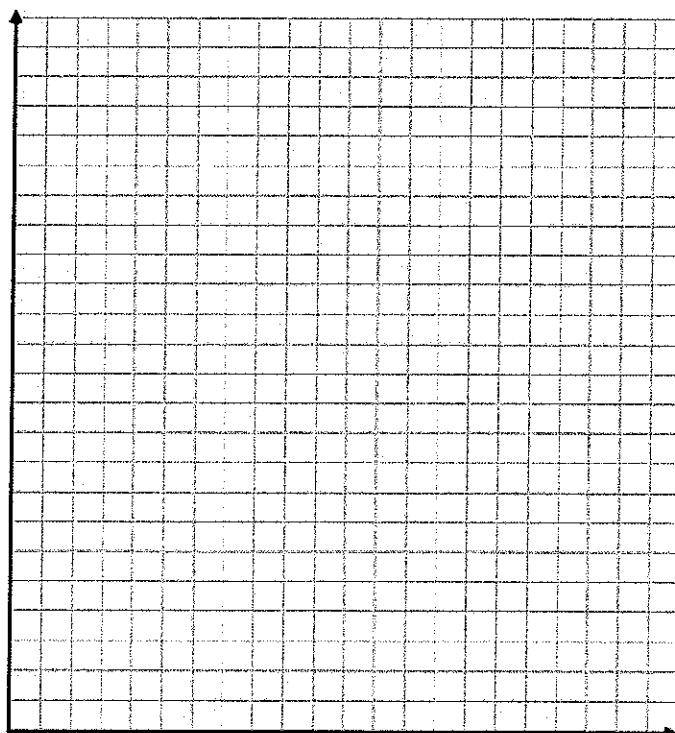
1. Professor Windley teaches a statistics class with 136 students. On the first day of class he records how much money each student has in his or her possession (in dollars) during the first class of the semester. The histogram displays the data. Describe the distribution.
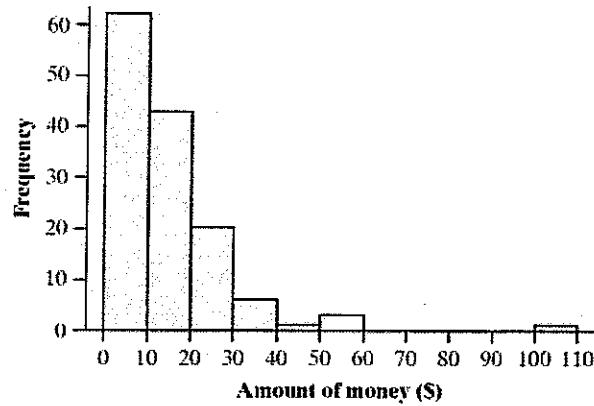
2. Researchers collected data on calories per serving for 77 brands of breakfast cereal. The histogram displays the data. Describe the distribution.
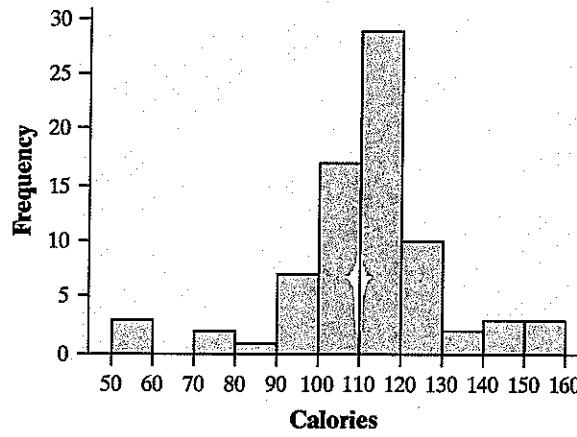
3. On August 7, 2007 Barry Bonds hit his 756th home run, breaking the all-time record, formerly held by Hank Aaron. Does that make Bonds a better home run hitter than Aaron? Let's compare their annual home run production over their entire careers. A side-by-side stemplot is shown below. Use the plot to compare the distributions of home runs per year for Bonds and Aaron.

Number of Home Runs Per Year

Bonds		Aaron
5	0	
9 6	1	0 2 3
8 6 5 5 4	2	0 4 6 7 9
7 7 4 4 3 3	3	0 2 4 4 8 9 9
9 6 6 5 5 2 0	4	0 0 4 4 4 4 5 7
	5	
	6	
3	7	
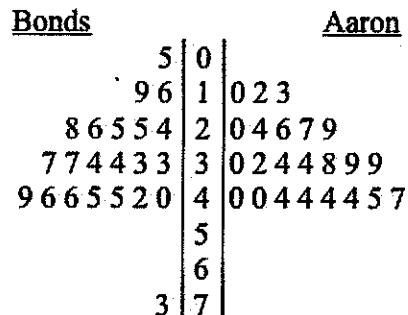
Key: 1|4 = 14 home runs

4. The histogram shows the heights of 300 randomly selected high school students. Which of the following is the best description of the shape of the distribution of heights?

a) Roughly symmetric and single-peaked (unimodal)
b) Roughly symmetric and double-peaked (bimodal)
c) Roughly symmetric and multi-peaked (multimodal)
d) Skewed to the left
e) Skewed to the right
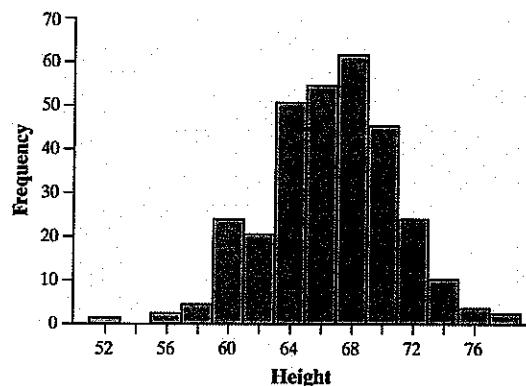
5. You look at real estate ads for houses in Naples, Florida. There are many houses ranging from $200,000 to $500,000 in price. The few houses on the water, however, are priced up to $15 million. The distribution of house prices will be

a) skewed to the left.
b) roughly symmetric.
c) skewed to the right.
d) single-peaked.
e) approximately uniform.

AP Statistics

Name: _____

Worksheet: Histograms Wkst

Directions: Create a frequency table and histogram for each set of data.

1. Chocolate candies per bag of trail mix, use 5 classes

50	42	100	45	68	32	100
67	61	31	75	39	62	64
49	55	51	33	99	96	64

Frequency table:

Interval	# of values
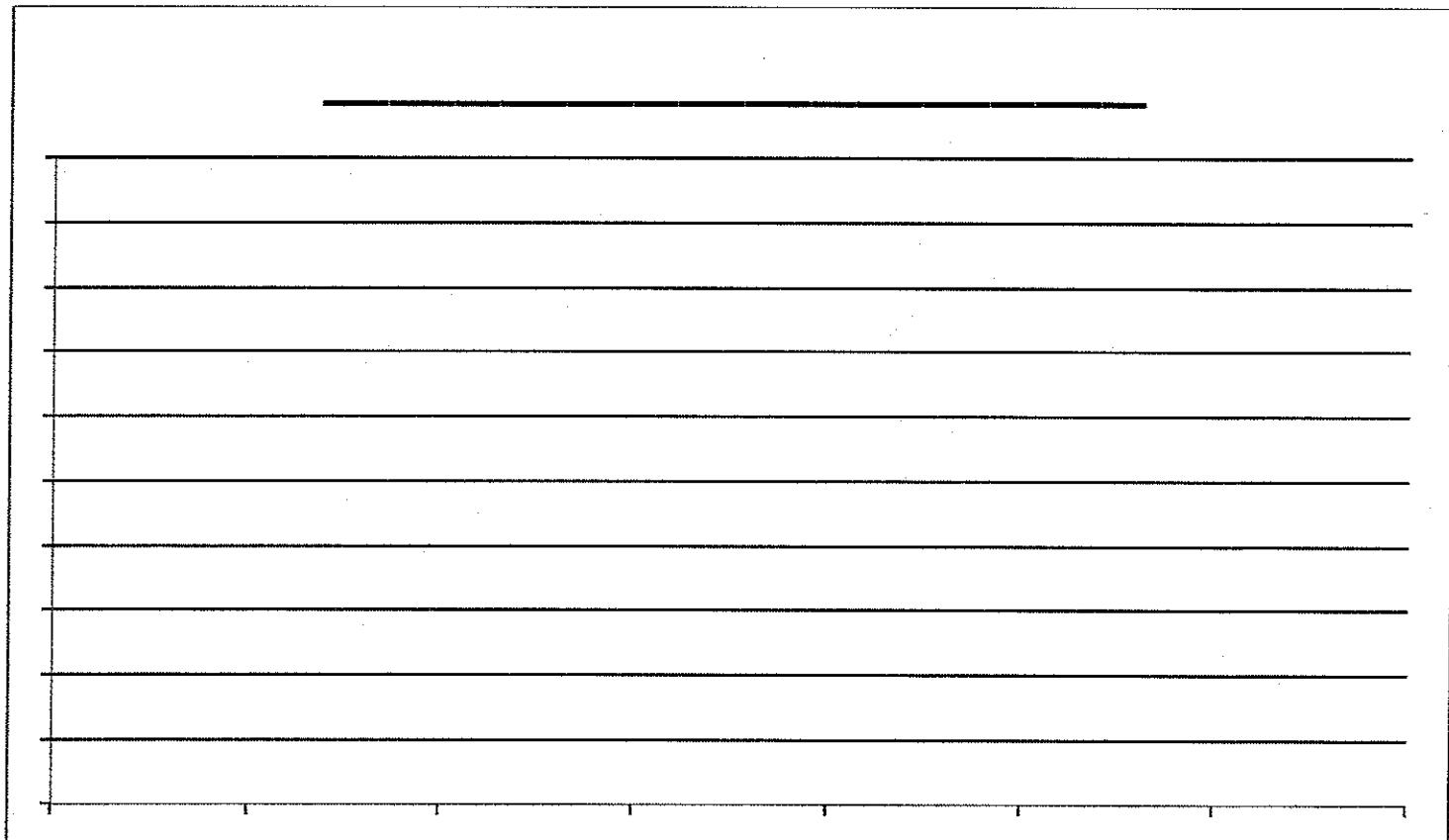
2. Test scores, out of 100 points, use 7 classes.

92	84	95	77	74	80	95	70	66
73	68	90	78	64	72	78	76	65
59	71	77	92	91	89	74	76	90

Frequency table:

Interval	# of values
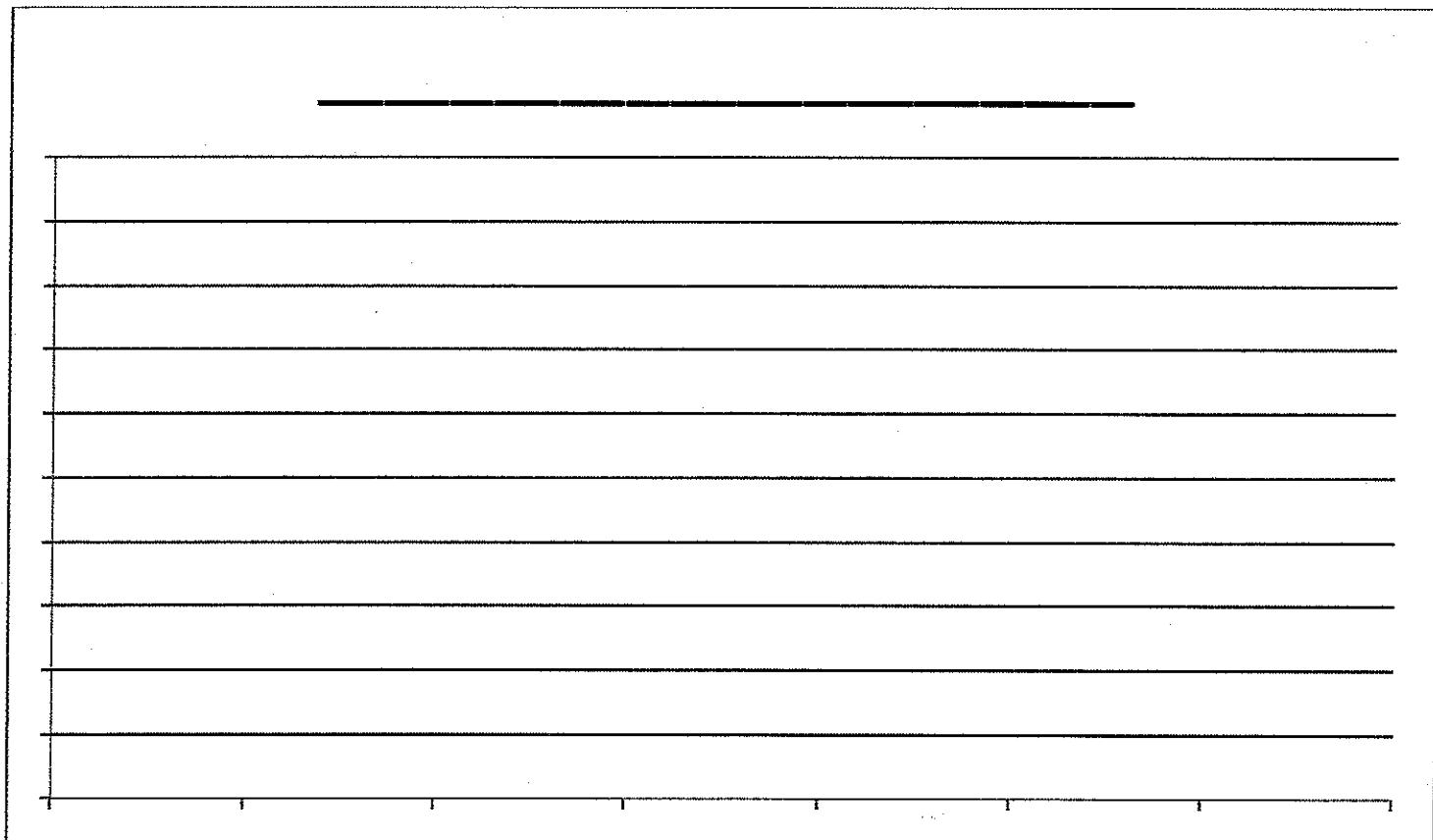
Directions: Answer the following questions based on each of the dot plots.

3. The dot plot below shows the number of students in each of the teacher's class.

Adam   Bara   Caelian   Daisy   Even   Dana   Dario   (classes)
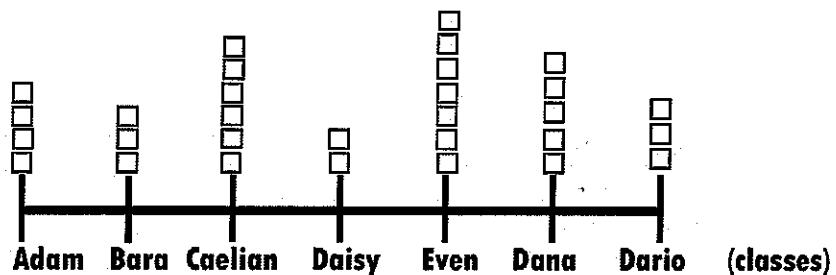
a. How many total students are there in all classes?

b. Which class has the least number of students?

c. Which class has the most number of students?

4. The dot plot shows the number of lemons each person has.

11   12   13   14   15   16   17   (lemons)
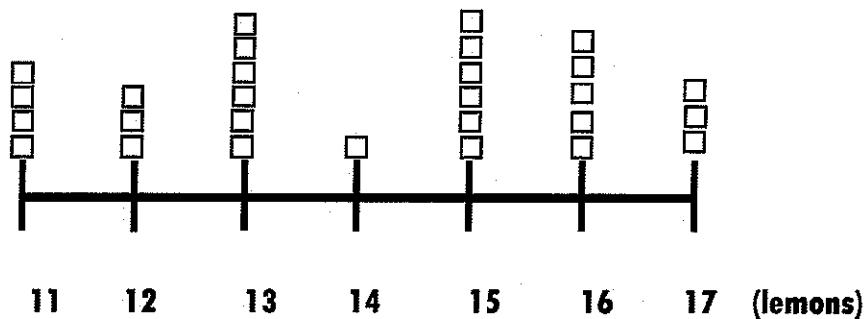
a. How many total individuals are represented in the dot plot?

b. What is the total number of lemons that the individuals have?

1. STATISTIC –WHAT IS THAT?

A statistic is a number calculated from data. Quantitative data has many different statistics that can be calculated.

a) Determine the given statistics from the data below on the number of homeruns Mark McGuire hit in each season from 1982 – 2001.

70	52	22	49	3	32	58	39
39	65	42	29	9	32	9	33

Mean	
Minimum	
Maximum	
Median	
Q1	
Q3	
Range	
IQR	

b) Construct a boxplot of the information from part a.
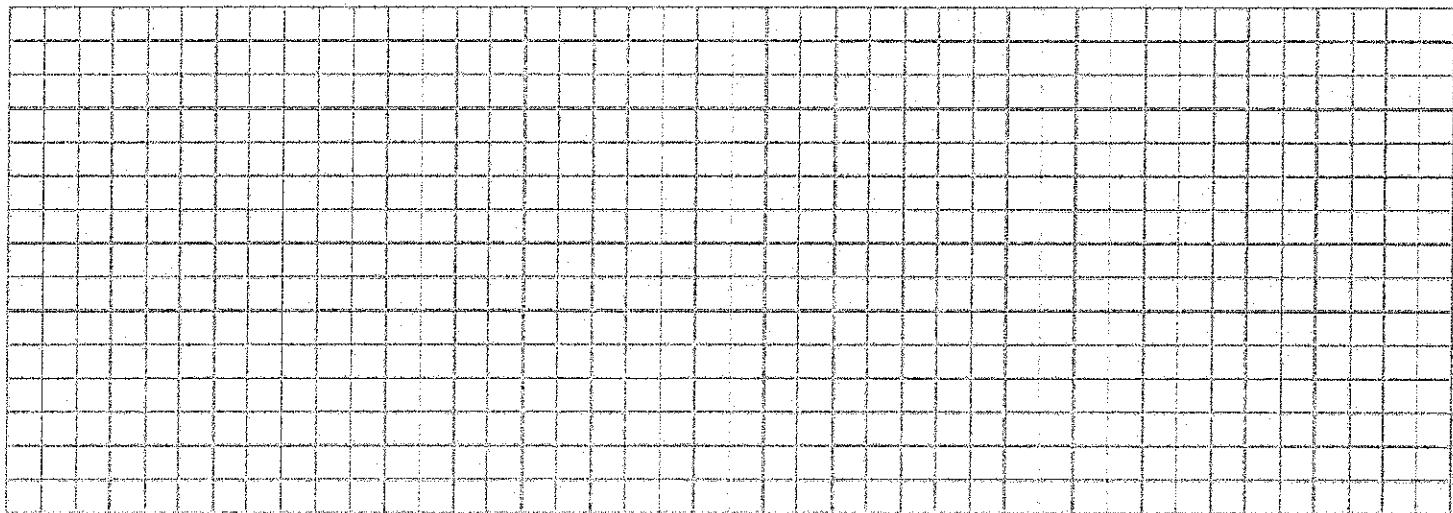
2. SSHA SCORES

Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

154 109 137 115 152 140 154 178 101 103 126 126 137 165 129 200 148 165

and for 20 first-year college men:

108 140 114 91 180 115 126 92 169 146 109 132 75 88 113 151 70 115 187 104

a) Use your calculator to find the summary statistics for both data sets.

Women			Men	
Mean			Mean	
Minimum			Minimum	
Q1			Q1	
Median			Median	
Q3			Q3	
Maximum			Maximum	
Range			Range	
IQR			IQR	

b) Using the minimum, Q1, Median, Q3, and Maximum from each gender, make parallel boxplots to compare the distributions.
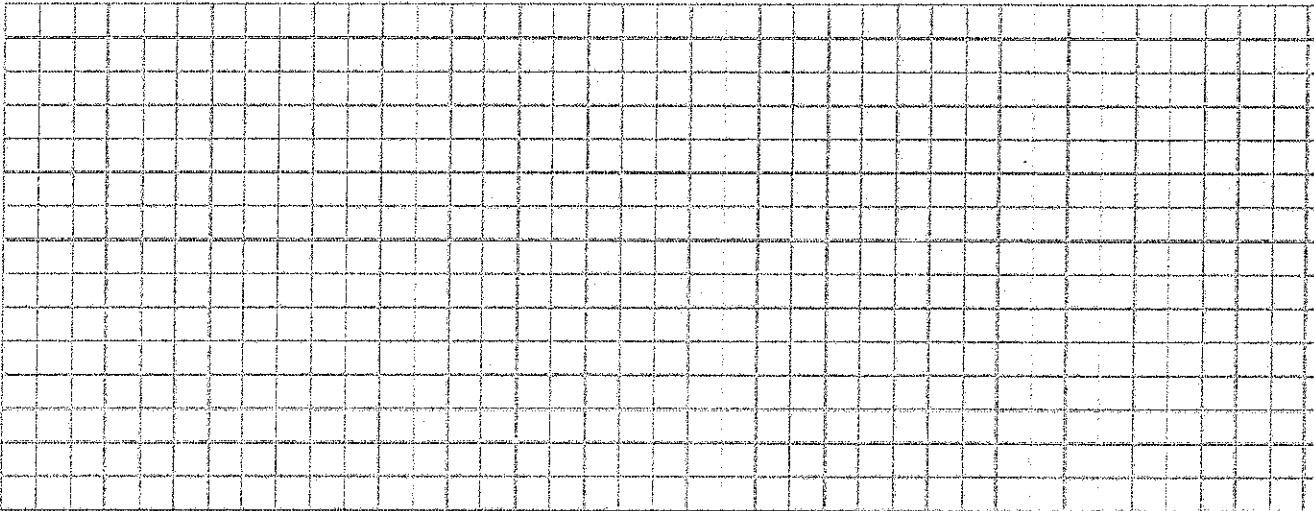
Questions 3 through 6 refer to the following:

The accompanying box-and-whisker plot represents the scores earned on a science test.

55 60 65 70 75 80 85 90 95 100
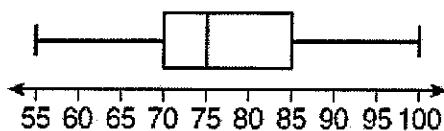
3. According to the diagram shown, what is the median score?

4. According to the diagram shown, what score represents the first quartile?

5. What statement is not true about the box and whisker plot shown?

A) 75 represents the mean score

B) 100 represents the maximum score

C) 85 represents the 3rd quartile

D) 55 represents the minimum score

6. A score of an 85 on the box-and-whisker plot shown refers to

A) the third quartile   B) the maximum score   C) the median   D) the mean

7. The median of any set of data always represents the

A) upper quartile   B) 50th percentile   C) mean of the data   D) 1st quartile

Questions 8 through 11 refer to the following:

The number of text messages 10 different students sent in 1 day is shown in the box-and-whisker plot below.

0 2 4 6 8 10 12 14 16 18 20 22
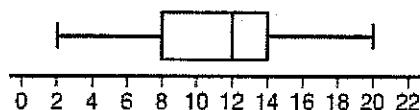
8. What is the minimum number of text messages sent according to the plot shown?

9. What number is at the 50th percentile according to the plot shown?

10. According to the plot shown, between what two numbers does half of the data lie?

11. According to the plot shown, how many text messages are at the 75th percentile?

12. For the following data, 43 54 55 63 67 68 69 77 85, suppose that the last data point is actually 115 instead of 85. What effect would this new maximum have on our value for the mean and our value for the median of the dataset?

13. Below you have a boxplot for the tar content of 25 different cigarettes. What is a plausible set of values for the five-number summary?

a) Min = 13, Q1 = 10, Median = 12.6, Q3 = 14, Max = 15
b) Min = 1, Q1 = 8.5, Median = 12.6, Q3 = 15, Max = 17
c) Min = 1, Q1 = 8.5, Median = 11.5, Q3 = 13, Max = 15
d) Min = 8.5, Q1 = 10, Median = 11.5, Q3 = 15, Max = 17
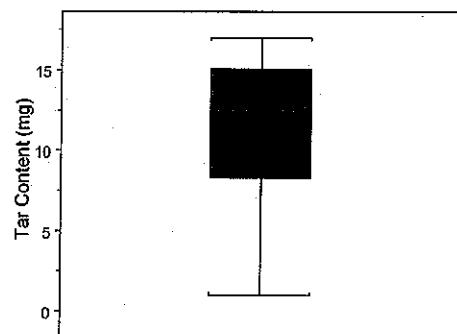
14. The shape of the boxplot in number 13 can be described as:

a. Bi-modal
c. Right-skewed
e. Uniform

b. Left-skewed
d. Symmetric

15. If a dataset contains outliers, which measure(s) of spread is resistant? Circle all that apply.

    a. Range   b. Interquartile range
    c. Standard deviation   d. Variance

16. Which of the following statements is TRUE?

    a. Standard deviation has no unit of measurement.
    b. Standard deviation is either positive or negative.
    c. Standard deviation is inflated by outliers.
    d. Standard deviation is used even when the mean is not an appropriate measure of center.

17. Look at the following side-by-side boxplots and compare the female and male shoulder girth.

    a. Females have a typically smaller shoulder girth than males.
    b. Females have a typically larger shoulder girth than males.
    c. Females and males have about the same shoulder girths.


Gender (0=Female, 1=Male)
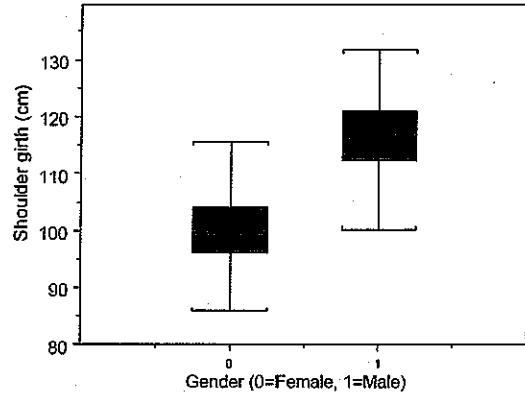
18. Look at the following side-by-side boxplots and compare the female and male thigh girth.

    a. Females have a typically smaller thigh girth than males.
    b. Females have a typically larger thigh girth than males.
    c. Females and males have about the same thigh girth.

Gender (0=Female, 1=Male)
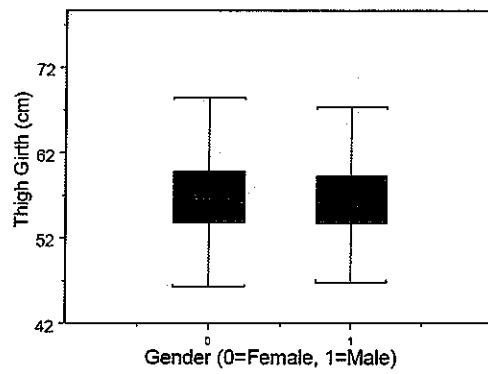
19. Compare the centers of Distribution A (Female Shoulder Girth) and Distribution B (Male Shoulder Girth) shown below.

Distribution A

Female Shoulder Girth (cm)

Distribution B

Male Shoulder Girth (cm)
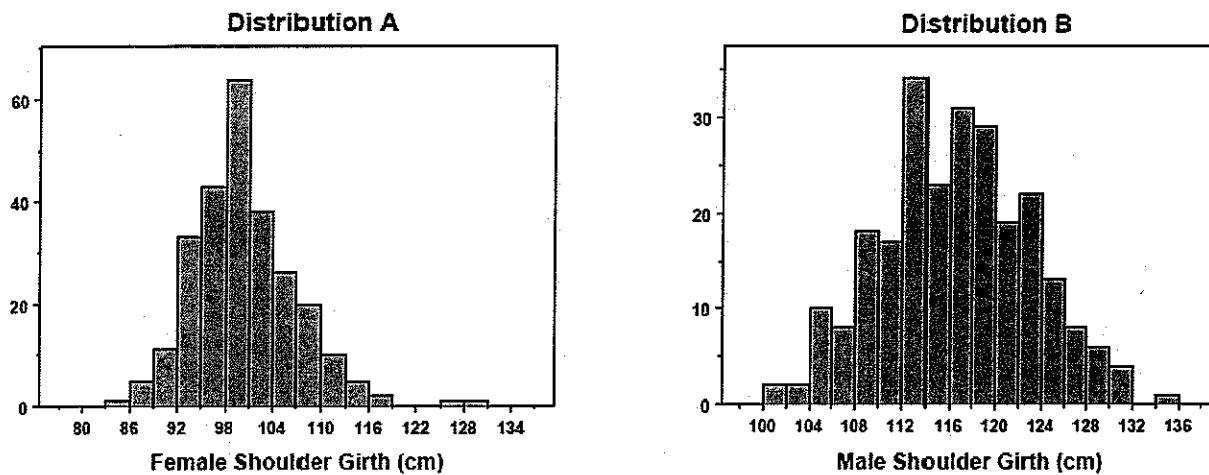
   a. The center of Distribution A is greater than the center of Distribution B.
   b. The center of Distribution A is less than the center of Distribution B.
   c. The center of Distribution A is equal to the center of Distribution B.

20. Compare the spreads of Distribution A (Female Shoulder Girth) and Distribution B (Male Shoulder Girth) from problem 8.

   a. The spread of Distribution A is greater than the spread of Distribution B.
   b. The spread of Distribution A is less than the spread of Distribution B.
   c. The spread of Distribution A is equal to the spread of Distribution B.

21. What is the approximate range and interquartile range of the Male Wrist Girth dataset shown below?

Male Wrist Girth (cm)