AP Statistics Summer Assignment

Welcome to AP Statistics! We are about to embark on a journey exploring data analysis. In order to be successful with this journey we need to begin by getting comfortable with vocabulary and the six primary data displays used during this Statistics’ class.

*The summer assignment is composed of two parts.*

1. **Reading and vocabulary:** You will use a free online Statistical tutoring site that will give you information on variables and data displays. While reviewing the information on the site you will be completing a vocabulary list. Complete the following steps to begin the tutorials.
   - Go to [www.stattrek.com](http://www.stattrek.com) We will be using this website and a few others throughout the course so get comfortable with using it!
   - Click on “AP Tutorial” and then scroll down to “Begin Tutorial”
   - On the left side of the screen is a list of general topics. When you move your cursor over them you will see subtopics. You will read the following subtopics to complete the vocabulary list. There are also short videos you can view for supplemental help.
   1. The Basics
      a. Variables
      b. Central Tendency
   2. Charts and Graphs
      a. Patterns in Data
      b. Dotplots
      c. Histograms
      d. Stemplots
      e. Boxplots
      f. Cumulative plot
      g. Scatter plot
      h. Comparing distributions
   3. Categorical Data
      a. One-way tables
   4. Surveys
      a. Sampling methods

2. **Practice Problems:** After reading all the material above and watching the short videos you should be able to complete the questions in the packet. You may do so in the spaces provided or on a separate piece of paper.
Upon your return to school in September, you are expected to turn in the completed packet. All work should have calculations and all data displays should be constructed neatly. This assignment will be graded for correctness. If you have any questions, please feel free to e-mail me at renika.johal@woodbridge.k12.nj.us. Have an amazing summer and remember be great everyday, even in the summer!

Ms. Johal
Part 1: Vocabulary List

1. Categorical Variables
   • Example:

2. Quantitative Variables
   • Example

3. Univariate Data

4. Bivariate Data

5. Median

6. Mean

7. Population
   • Example:

8. Sample
   • Example:

9. Center

10. Spread
11. Symmetry

12. Unimodal and Bimodal

13. Skewness
   
   - Sketch Skewed Left:
   
   - Sketch Skewed Right:

14. Uniform

15. Gaps

16. Outliers

17. Dotplots

18. Difference between bar chart and histogram

19. Stemplots
20. Boxplots

21. Quartiles

22. Range

23. Interquartile Range

24. Parallel boxplots

25. Parameter

26. Statistic
Part 2: Practice Problems

Determine if the variables listed below are QUANTITATIVE or CATEGORICAL.

1. Time it takes to get to school
2. Number of people under 18 living in a household
3. Hair color
4. Temperature of a cup of coffee
5. Teacher salaries
6. Gender
7. Smoking
8. Height
9. Amount of oil spilled
10. Age of Oscar winners
11. Type of depression medication
12. Jellybean flavors
13. Country of origin
14. Type of meat
15. Number of shoes owned

A statistic is a number calculated from data. You have seen statistics all around you when listening to the news, watching sports, and among other daily routines in your life. Quantitative data has many different statistics that can be calculated. Determine the given statistics from the data below on the number of three pointers LeBron James hit in each season of his professional NBA career.

<table>
<thead>
<tr>
<th>63</th>
<th>108</th>
<th>127</th>
<th>99</th>
<th>113</th>
<th>132</th>
<th>129</th>
</tr>
</thead>
<tbody>
<tr>
<td>92</td>
<td>54</td>
<td>103</td>
<td>116</td>
<td>120</td>
<td>87</td>
<td></td>
</tr>
</tbody>
</table>

Mean = 

Minimum = 

Maximum = 

Median = 

Q₁ = 

Q₃ = 

Range = 

IQR =
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, 4,051 from drowning, and 3,601 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.

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. You may try using Excel or you can create one by hand.
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.

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>1</th>
<th>4</th>
<th>3</th>
<th>7</th>
<th>2</th>
<th>3</th>
<th>3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A. Make a dotplot to display these data. Make sure you include appropriate labels, title, and scale.
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 (rounded to the nearest dollar), arranged in increasing order.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>26</td>
<td>26</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>36</td>
<td>39</td>
<td>39</td>
<td>41</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>45</td>
<td>47</td>
<td>49</td>
</tr>
<tr>
<td>50</td>
<td>53</td>
<td>55</td>
<td>59</td>
<td>61</td>
<td>70</td>
<td>83</td>
<td>86</td>
<td>86</td>
<td>93</td>
</tr>
</tbody>
</table>

A. Make a stemplot using tens of dollars as the stem and dollars as the leaves. Make sure you include appropriate labels, title, and key.
WHERE DO OLDER FOLKS LIVE?

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

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

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.

A. 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.

<table>
<thead>
<tr>
<th>Bin Widths</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 to &lt; 6</td>
<td>1</td>
</tr>
<tr>
<td>6 to &lt; 8</td>
<td></td>
</tr>
<tr>
<td>8 to &lt; 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SSHA SCORES

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

<table>
<thead>
<tr>
<th>154</th>
<th>109</th>
<th>137</th>
<th>115</th>
<th>152</th>
<th>140</th>
<th>154</th>
<th>178</th>
<th>101</th>
</tr>
</thead>
<tbody>
<tr>
<td>103</td>
<td>126</td>
<td>126</td>
<td>137</td>
<td>165</td>
<td>165</td>
<td>129</td>
<td>200</td>
<td>148</td>
</tr>
</tbody>
</table>

And for 20 first year college men:

<table>
<thead>
<tr>
<th>108</th>
<th>140</th>
<th>114</th>
<th>91</th>
<th>180</th>
<th>115</th>
<th>126</th>
<th>92</th>
<th>169</th>
<th>146</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
<td>132</td>
<td>75</td>
<td>88</td>
<td>113</td>
<td>151</td>
<td>70</td>
<td>115</td>
<td>187</td>
<td>104</td>
</tr>
</tbody>
</table>

A. Put the data values in order for each gender. Compute numeral summaries for each gender.

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th></th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td></td>
<td>Minimum</td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td></td>
<td>Q1</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td>Median</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td>Q3</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td></td>
<td>Maximum</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>Range</td>
<td></td>
</tr>
<tr>
<td>IQR</td>
<td></td>
<td>IQR</td>
<td></td>
</tr>
</tbody>
</table>

B. Using the minimum, Q1, Median, Q3, and Maximum from each gender, make parallel boxplots to compare the distribution.
ALGEBRA PAGE!

The prerequisite for AP Statistics is Algebra II. You will not find very much equation solving in this course, but some quick review of Algebra I and Algebra II content will be helpful.

*In Algebra I you learned about methods of sampling design. For more review, refer back to your reading on www.Stattrek.com “Survey Sampling Methods”.*

The 7 types of sampling designs are:

A. Voluntary response  
B. Convenience  
C. Simple Random  
D. Stratified  
E. Cluster  
F. Multistage  
G. Systematic

1. The Maryland division of Weight Watchers is doing research to determine how many people on the Weight Watchers diet cheat at least once a week. They decide that anonymous surveys will give them an accurate representation but do not have time to get responses from ALL the Maryland Weight Watchers people.

*Read the scenarios belows and determine which of the 7 sampling methods best describes it.*

__________ I. Randomly select 10 members from each of the WW centers in the Maryland division.

__________ II. Use an alphabetical listing of all Maryland division members. Randomly choose a starting person on the list. Then select every 20th person thereafter.

__________ III. Randomly select 2 or 3 branches of the Maryland division and survey every member of that center.

__________ IV. Send out the survey to every member of the Maryland division. Place drop boxes in each WW center. Anyone who returns a survey will be in the sample.

__________ V. The Maryland regional office is in Baltimore so they survey members at the WW center in Baltimore.

__________ VI. From a numbered list of all Maryland division members, use a computer to randomly select 100 numbers and survey all members with those
corresponding numbers.

2. What is the population of interest in the WW situation?

Here is a formula that is used often in AP Statistics: \( z = \frac{x - \overline{x}}{s} \).

1. If \( z = 2.5 \), \( x = 102 \), and \( \overline{x} = 100 \), what is \( s \)? Show your work.

2. If \( z = -3.35 \), \( x = 60 \), and \( s = 4 \), what is \( \overline{x} \)? Show your work.
It is expected that you have a thorough understanding of linear functions.

1. The USDA reported that in 1990 each person in the United States consumed an average of 133 pounds of natural sweeteners. They also claim this amount has decreased by about 0.6 pounds each year.
   a. Write a linear equation that relates years since 1990 to the average consumption of natural sweeteners. Define your variables.

   b. What is the slope and what is the y-intercept?

   c. Predict the average consumption of sweeteners per person for the year 2005.

2. The following equation can be used to predict the average height of boys anywhere between birth and 15 years old: \( y = 2.79x + 25.64 \), where \( x \) is the age (in years) and \( y \) is the height (in inches).
   a. What does the slope represent in this problem? Interpret it in context.

   b. What does the y-intercept represent in this problem? Interpret it in context.
You are expected to have a basic understanding of simple probability.

1. A special lottery is to be held to select the student who will live in the only deluxe room in a dormitory. There are 100 seniors, 150 juniors, and 200 sophomores who applied. Each senior’s name is placed in the lottery 3 times; each junior’s name, 2 times; and each sophomore’s name, 1 time. What is the probability that a senior’s name will be chosen?

   A. $\frac{1}{8}$   B. $\frac{2}{9}$   C. $\frac{2}{7}$   D. $\frac{3}{8}$   E. $\frac{1}{2}$

2. Which of the following has a probability closest to 0.5?

   A. The sun will rise tomorrow
   B. It will rain tomorrow
   C. You will see a dog with only three legs when you leave the room
   D. A fair die will come up with a score of 6, four times in a row
   E. There will be a plane crash somewhere in the world within the next 5 minutes.

3. If a coin is tossed twice, what is the probability that on the first toss the coin lands on heads and on the second toss the coin lands on tails?

   A. $\frac{1}{6}$   B. $\frac{1}{3}$   C. $\frac{1}{4}$   D. $\frac{1}{2}$   E. 1

4. If a coin is tossed twice what is the probability that it will land either on heads both times or tails both times?

   A. $\frac{1}{8}$   B. $\frac{1}{6}$   C. $\frac{1}{4}$   D. $\frac{1}{2}$   E. 1

5. Calculate the following probabilities and arrange them in order from least to greatest.

   I. The probability that a fair die will produce an even number.

   II. A random digit from 1 to 9 (inclusive) is chosen, with all digits being equally likely. The probability that when it’s squared it will end with the digit 1.

   III. The probability that a letter chosen from the alphabet will be a vowel.

   IV. A random number between 1 and 20 (inclusive) is chosen. The probability that its square root will not be an integer.

   ORDER______________________________