To incoming AP Statistics students,
To be best prepared for your AP Statistics course, you will complete this packet, which you will submit on the first day of the first full week of school. This assignment will be graded and count as the first assignment of the first marking period. You will be able to access this assignment through ltps.org and/or classroom.google.com. The code you will need for google classroom is: 7kbcdr

Google classroom Code:
- Go to www.classroom.google.com
- Sign into your school provided .info account.
- Enter access code 7kbcdr

This Google group has been established in order to provide you, the students, with support as a group. If you are struggling with concepts/material, there is a blog option to post questions to one another. In addition, you have the educational resources listed below for assistance. Remember, the math course from the current school year is the prerequisite course for the course you have enrolled into for the Fall. Your personal notebook and handouts from this year’s class is a resource that is at your disposal. The instructor may monitor the Google Classroom throughout the summer. If you need the instructor’s help, you will still use the Google classroom to reach out.

Directions:
- Complete all work
- Work should be organized and neatly labeled. If you use separate pages to show your work, be sure each additional work page has your name on it.
- Any problem with no work shown will receive 0 points.
- When asked to explain or justify, you must write in full sentences.

Scoring/Grading: The AP Statistics Summer Assignment is worth a total of 45 points.
- Question #1 is worth 6 pts
- Question #2 is worth 5 pts
- Question #3 is worth 10 pts
- The summary statistics and two graphing data sections are each worth 8 points.

This will count as your first grade. The assignment is due on the first day of the first full week of school.

Resources:
For additional examples and support you can reference any of the sites listed below and search the skill/concept.

- stattrek.com
- KhanAcademy.com
- You Tube.com or Teacher Tube.com
- MathIsPower4u.com
- IXL.com

See you in September,
Mr. Krajunus
1. In a statistics class there are 18 juniors and 10 seniors; 5 of the seniors are females, and 10 of the juniors are males. Complete the table below:

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If a student is selected at random, find the probability of selecting

(a) a junior or a female

(b) not a junior male

2. Consolidated Builders has bid on two large construction contracts. The company president believes that the probability of winning the first contract (event A) is 0.5, that the probability of winning a second (event B) is 0.3, and that the probability of winning both jobs is 0.1.

<table>
<thead>
<tr>
<th></th>
<th>Winning second bid</th>
<th>Losing second bid</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winning first bid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losing first bid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Construct a Venn diagram that summarizes what you know about events A and B.
(b) What is the probability that Consolidated wins at least one of the job?

(c) Use the information above to calculate the probability of each.

i. Consolidated wins both jobs.

ii. Consolidated wins the first job but not the second.

iii. Either Consolidated does not win the first job or wins the second.

iv. Consolidated does not win either job.

3. Consider the following activity: The letters in the word AARDVARK are printed on identical plastic cards with one letter per card. The eight cards are then placed in a hat, and one card is randomly chosen (without looking) from the hat. The chance process we are interested in is what letter is on the selected card.

(a) List the sample space S of all possible outcomes.

(b) Make a table that shows the set of outcomes and the probability of each outcome:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
</tr>
</thead>
</table>

(c) Consider the following events:

V: the letter chosen is a vowel.
F: the letter chosen falls in the first half of the alphabet (that is, between A and M).

List the outcomes in each of the following events, and determine their probabilities:

\[ F^c \text{ (not } F) = \{ \]  
\[ P(\text{Not } F) = \]

\[ V \text{ and } F = \{ \]  
\[ P(V \text{ and } F) = \]

\[ V \text{ given } F = \{ \]  
\[ P(V \text{ given } F) = \]
(d) Are the events \( V \) and \( F \) are independent? Explain.

(e) Are the events \( V \) and \( F \) mutually exclusive? Explain.

(f) Name one famous AARDVARK

4. Faust Intermediate School surveys the families of its students and determines the following: if a family is chosen at random, the probability that they own a dog is 0.38, the probability they own a cat is 0.23, and the probability they own both a dog and a cat is 0.12.

(a) Let \( D \) = randomly-chosen family owns a dog, and \( C \) = randomly-chosen family owns a cat. Complete the Venn diagram that summarizes the probabilities above.

(b) Find each of the following.
   i. The probability that a randomly-selected family owns a dog or a cat.

   ii. The probability that a randomly-selected family owns a dog or doesn’t own a cat.

   iii. The probability that a randomly-selected family doesn’t own a dog and doesn’t own a cat.
Summary Statistics – Center and Spread

A statistic is a number that gives information about a set of data. Common examples include mean, median, mode (which we won’t worry about in AP Stat), range, standard deviation, and more!

SYMBOLS
In statistics, we use a variety of symbols to represent statistics. Sometimes, the symbol used depends on whether we are talking about a population or a sample (select members of a given population).

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Standard Deviation</th>
<th>Median</th>
<th>Number of data points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>(\mu) (&quot;mu&quot;)</td>
<td>(\sigma) (&quot;sigma&quot;)</td>
<td>No symbol</td>
<td>(n)</td>
</tr>
<tr>
<td>Sample</td>
<td>(\overline{x}) (&quot;x-bar&quot;)</td>
<td>(s)</td>
<td>(Often abbreviated “Med.”)</td>
<td></td>
</tr>
</tbody>
</table>

Measures of CENTER
The center of a data set lets us understand the “average” or “typical” value of a number in that data set. There are two main measures of center: mean and median.

**MEAN**
Add up all data points, then divide by the number of data points.
- \(\mu\) (or \(\overline{x}\)) = \(\frac{\sum x}{n}\)
- “Sum of data points over number of data points”

**Example 1:** Science grades of a sample of 15 juniors:
91, 87, 66, 74, 85, 98, 43, 88, 77, 62, 83, 91, 89, 52, 100

This is a SAMPLE, so \(\overline{x} = \frac{\sum x}{n} = \frac{1186}{15} = 79.07\)

**Example 2:** Heights of all 6 people in a family (inches):
47, 58, 61, 65, 68, 70

This is the POPULATION, so \(\mu = \frac{\sum x}{n} = \frac{369}{6} = 61.5\)

**MEDIAN**
The middle number of the data set, assuming that the data points are in order (smallest to largest)
- If there are 2 numbers in the middle, find the mean of those two numbers!
- A nice trick for finding the position of the median is to use \(\frac{n+1}{2}\)

**Example 1:** Science grades of a sample of 15 juniors:
91, 87, 66, 74, 85, 98, 43, 88, 77, 62, 83, 91, 89, 52, 100

\(\frac{n+1}{2} = \frac{15+1}{2} = 8\). Median is the 8th number (IN ORDER)
43, 52, 62, 66, 74, 77, 83, 85, 87, 88, 91, 91, 98, 100

**Example 2:** Heights of all 6 people in a family (inches):
47, 58, 61, 65, 68, 70

\(\frac{n+1}{2} = \frac{6+1}{2} = 3.5\). Median is between the 3rd & 4th number
47, 58, 61, 65, 68, 70; Average = \(\frac{61+65}{2} = 63\)

Measures of SPREAD
The spread of a data set tells us whether the data points are far apart or clustered together. The most important measure of spread is standard deviation, which is the typical distance of the data points from the mean. Other measures of spread, such as Range and IQR, will be discussed in Part 5.

The formulas for Standard Deviation are as follows. Note that they are slightly different for a population and a sample (the sample one will be slightly larger to account for the fact that the sample doesn’t include all members of a population)

Population: \(\sigma = \sqrt{\frac{\Sigma (x_i - \mu)^2}{n}}\)
Sample: \(s = \sqrt{\frac{\Sigma (x_i - \overline{x})^2}{n-1}}\)
*You will NOT have to calculate Standard Deviation by hand in this course!*  
What you will have to do, however, is be able to interpret and compare the Standard Deviations of different data sets:

- **Larger Standard Deviation:** The data is more spread out (points are typically further from the mean)
- **Smaller Standard Deviation:** The data is closer together (points are typically closer to the mean)

**Example:**

Data Set 1: 1, 2, 3, 17, 18, 19; \( \mu = 10, \sigma = 8.04 \)  
Data Set 2: 7, 8, 9, 11, 12, 13; \( \mu = 10, \sigma = 2.16 \)

Notice how Data Set 1 is more spread out, while Data Set 2 is closer together. This is reflected in the fact that Set 1’s Standard Deviation (8.04) is higher than Set 2’s Standard Deviation (2.16)

**Practice Problems**

1. Find the mean and median of the following data set. **Show work** when appropriate!
   
   Teaching experience of all LHS math teachers (n = 15): 1, 3, 3, 3, 4, 4, 5, 5, 5, 6, 7, 7, 18, 23, 26
   
   Symbol for mean: _______ Value of mean: _______ Position of Median: _______ Value of Median: _______

2. Find the mean and median of the following data set. **Show work** when appropriate!
   
   Weights of 8 randomly-selected chickens on a farm (in pounds): 5.4, 5.7, 6.2, 6.9, 7.2, 7.2, 8.1, 9.0
   
   Symbol for mean: _______ Value of mean: _______ Position of Median: _______ Value of Median: _______

3. Find the mean and median of the following data set. **Show work** when appropriate!
   
   Temperature readings on all thermostats in an office building: 71, 72, 72, 74, 68, 74, 71, 72, 69, 76
   
   Symbol for mean: _______ Value of mean: _______ Position of Median: _______ Value of Median: _______

4. List the following data sets in order from least spread out to most spread out. Then, write 1-2 sentences explaining how you could tell.
   
   LHS Math Teachers: \( \sigma = 7.02 \)  
   Chickens: \( s = 1.21 \)  
   Thermostats: \( \sigma = 2.26 \)
Graphing Data – Dotplots and Stem-and-Leaf Plots

Statistics such as mean, median, and standard deviation are very useful in summarizing data and giving overall trends. But they don’t tell the full story. By making a graph of the data, we can go beyond the numbers and see shapes and patterns in the data. Shown below are two common ways in which to graph data.

**Dotplots**
- Make an AXIS on the bottom (you can go by 1s, 2s, 5s, 10s…whatever makes sense for the data!)
- Put one dot for each data point on the axis. If there is more than one data point for a given value, stack the dots!

**Example:** Number of matches in 20 randomly-selected boxes.
45, 47, 48, 49, 49, 49, 50, 50, 50, 50, 50, 50, 50, 50, 50, 51, 51, 51, 52, 52

![Graph of matches per box](image)

**Example:** Weights of players on a high school baseball team
140, 140, 143, 144, 144, 148, 148, 148, 148, 152, 152, 152, 156, 156, 156, 156, 160, 164, 164, 168, 168

![Graph of weights](image)

**Stem-and-Leaf Plots (also called Stemplots)**
- Use a KEY to determine what the stems and leaves are worth
- DO NOT SKIP STEMS. If there are no data points for that stem, just keep the stem there and put no leaves after it. Skipping the stem will alter what the stemplot looks like.

**Example:** Temperatures at OU football games, 2009
95, 101, 86, 82, 70, 74, 63, 72, 63

**Example:** Gross National Product (per capita) of West African countries
180, 240, 260, 270, 310, 330, 360, 370, 390, 410, 480, 500, 710, 730, 890

![Stem-and-Leaf Plot](image)

**Practice Problems**

Fuel Economy for a Random Sample of 2015 Model Year Vehicles
1. List all the data points in the dotplot:

16 18 20 22 24 26 28 30 32 34 36 38

2. Using the dotplot shown, find the **mean** and **median** 100-meter sprint time

![Dotplot](image)

3. The following stemplot shows the final exam scores of a class with 10 students. List the scores of each student:

<table>
<thead>
<tr>
<th>stem</th>
<th>leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7    8 8 9</td>
</tr>
<tr>
<td>9</td>
<td>0    6 7 7</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Key: 6|8 means 68

4. Find the **mean** and **median** of the data shown in the stemplot.  
**NOTE:** Look at the key carefully!

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0    2 3 6</td>
</tr>
<tr>
<td>3</td>
<td>2    3 5 6 7</td>
</tr>
<tr>
<td>4</td>
<td>6    8 9</td>
</tr>
<tr>
<td>5</td>
<td>4    7</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

**KEY:** 4|6 = 4.6
Graphing Data – Box and Whisker Plots

Example: Ages of 9 employees in an office

\[37, 24, 51, 46, 62, 28, 35, 49, 55\]

*Finding MEDIAN:* (9 people. \(\frac{9+1}{2} = 5\), it’s the 5\(^{th}\) number. Remember that numbers must be in order!)

\[24, 28, 35, 37, 46, 49, 51, 55, 62\]

Median = 46

*Finding QUARTILES:* There are 9 data points, so split the data set in half! **NOTE:** If there is one number that serves as the median, as with this data set, it is not included in either half!

\begin{align*}
\text{Q1} & : 24, 28, 35, 37, 46 (\text{not included}) \\
\text{Average: } & 31.5 \\
\text{Q3} & : 49, 51, 55, 62 \\
\text{Average: } & 53
\end{align*}

*BOXPLOT:* Min = 24, Q1 = 31.5, Med = 46, Q3 = 53, Max = 62

RANGE: 62 – 24 = 38

IQR: 53 – 31.5 = 21.5

Practice Problems
1. Construct a box and whisker plot for the following data set. Be sure to include an axis like the ones in the examples!

\[17, 21, 24, 26, 31, 33, 36, 37, 41, 48\]