Dear Intro to Statistics Student,

Congratulations on your decision to join the thousands of other students across the country that will be enrolled in Intro to Statistics in the upcoming school year. You have joined the growing ranks of students who recognize the need to take an introductory statistics course. A course such as this is typically required of many college majors including the social sciences, health sciences, and business. Since statistics spans all subject areas, you will also need to possess common and, at times, uncommon knowledge of other areas of everyday life, such as geography, politics, science, sports and human nature. The best way to attain this knowledge is to acquire information from as many resources as possible. Watch the news every night; read newspapers or websites pertaining to these and other topics. Nothing is off topic when it comes to Statistics.

Statistics is not an easy class. You can expect to spend time studying outside of class, as well as in class. However, Statistics is special. It is a course that combines both mathematical and verbal skills. On our exams, you will be asked to write descriptive paragraphs and concluding sentences. You will have to explain the reasoning behind the method you use and the conclusions you find.

Throughout the course of the year, Statistics will expose you to four themes: exploring data, sampling and experimentation, anticipation patterns, and statistical inference. Statistics describes the qualitative world around us in a quantitative manner. A Statistics class insists that you learn how to describe those quantities in a qualitative manner as well. Everything is connected!

How can you prepare for this class? Complete the summer assignment by the first day of class. It will be collected and graded for completion and effort. You are expected to use the internet to research the following vocabulary, some of which you should be familiar with: categorical variable, quantitative variable, bar graph, dot plot, survey, experiment, observational study, measures of central tendency, and measures of dispersion.

You should also get the required supplies for the course which includes graph paper, notebook, pencil, and a TI-83 or TI-84.

If you have any questions, do not hesitate to e-mail me over the summer at dborroni@lawrencewoodmere.org Again, welcome to Intro to Statistics!

Good luck and I look forward to seeing you in September.

Sincerely,

Ms. Danielle Borroni
dborroni@lawrencewoodmere.org
This assignment is due on the FIRST DAY OF SCHOOL. The assignment is mandatory, there are NO exceptions. The questions MUST be answered on a separate sheet of paper (graph paper may be used if desirable). To assist with the assignment, you are expected to use the book and the internet to research the following vocabulary, some of which you should already be familiar with:

- Categorical Variables
- Quantitative Variables
- Bar Graph
- Dotplot
- Survey
- Experiment
- Observational Study
- Measures of Central Tendency
- Measures of Dispersion

WHY STATISTICS?

Write a page explaining why high school students should take a statistics class. First, use evidence from the following sources to make your case:


Then, write a paragraph explaining what you hope to gain from taking a class in Statistics. What are your reasons for signing up for this class?

STUDYING VARIABLES

1. Here is a small part of a data set that describes Major League Baseball players as of opening day of the 2005 season:

<table>
<thead>
<tr>
<th>Player</th>
<th>Team</th>
<th>Position</th>
<th>Age</th>
<th>Height</th>
<th>Weight</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ortiz, David</td>
<td>Red Sox</td>
<td>Outfielder</td>
<td>29</td>
<td>6-4</td>
<td>230</td>
<td>5,250,000</td>
</tr>
<tr>
<td>Nix, Laynce</td>
<td>Rangers</td>
<td>Outfielder</td>
<td>24</td>
<td>6-0</td>
<td>200</td>
<td>316,000</td>
</tr>
<tr>
<td>Perez, Antonio</td>
<td>Dodgers</td>
<td>Infielder</td>
<td>25</td>
<td>5-11</td>
<td>175</td>
<td>320,000</td>
</tr>
<tr>
<td>Piazza, Mike</td>
<td>Mets</td>
<td>Catcher</td>
<td>36</td>
<td>6-3</td>
<td>215</td>
<td>16,071,429</td>
</tr>
<tr>
<td>Rolen, Scott</td>
<td>Cardinals</td>
<td>Infielder</td>
<td>30</td>
<td>6-4</td>
<td>240</td>
<td>10,715,509</td>
</tr>
</tbody>
</table>

   a. What individuals does this data set describe?
   b. Identify the variables that were recorded. Label each as categorical or quantitative.
   c. Based on the data in the table, what do you think are the units of measurement for each of the quantitative variables?

2. How can we help wood surfaces resist weathering, especially when restoring historic wooden buildings? A study of this question prepared wooden panels and then exposed them to the weather. Here are some of the variables recorded. Identify each of these variables as categorical or quantitative. Then give a possible value for each variable.

<table>
<thead>
<tr>
<th>Categorical or Quantitative</th>
<th>Possible Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Type of wood</td>
<td></td>
</tr>
<tr>
<td>b. Water repellent</td>
<td></td>
</tr>
<tr>
<td>c. Paint thickness</td>
<td></td>
</tr>
<tr>
<td>d. Paint color</td>
<td></td>
</tr>
<tr>
<td>e. Weathering time</td>
<td></td>
</tr>
</tbody>
</table>
3. Popular Magazines often rank cities in terms of how desirable it is to live and work there.
   a. Identify two categorical variables and two quantitative variables that could be used to measure a city's desirability. Give a reason for each of your choices.
   b. In what units would you measure each quantitative variable?

4. Categorical or Quantitative Data
   a. Number of students in a classroom
   b. Sex of students in a classroom
   c. Height of students in the classroom
   d. Temperature in the classroom
   e. Age
   f. Smoke or Not
   g. Republican, Democrat, or Neither
   h. Colors of the Rainbow
   i. Grade you are in
   j. Whether you are a freshman, sophomore, junior, or senior

REPRESENTING CATEGORICAL DATA

5. Births are not, as you might think, evenly distributed across the days of the week. Here are the average numbers of babies born on each day of the week is 2003.
   a. Is the variable shown in the chart categorical or quantitative? Why?
   b. Present these data in a well-labeled BAR GRAPH.
   c. Suggest some possible reasons why there are fewer births on weekends.

<table>
<thead>
<tr>
<th>Day</th>
<th>Births</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunday</td>
<td>7,563</td>
</tr>
<tr>
<td>Monday</td>
<td>11,733</td>
</tr>
<tr>
<td>Tuesday</td>
<td>13,001</td>
</tr>
<tr>
<td>Wednesday</td>
<td>12,598</td>
</tr>
<tr>
<td>Thursday</td>
<td>12,514</td>
</tr>
<tr>
<td>Friday</td>
<td>12,396</td>
</tr>
<tr>
<td>Saturday</td>
<td>8,605</td>
</tr>
</tbody>
</table>

6. News from the auto color front: fewer luxury car buyers are choosing "neutral colors" (Silver, white, black). Here is the distribution of the most popular colors for 2005 model luxury cars made in North America.
   a. What percent of vehicles are some other color not listed in the chart?
   b. Make a bar graph of the color data.

<table>
<thead>
<tr>
<th>Color</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver</td>
<td>20%</td>
</tr>
<tr>
<td>White, Pearl</td>
<td>18%</td>
</tr>
<tr>
<td>Black</td>
<td>16%</td>
</tr>
<tr>
<td>Blue</td>
<td>13%</td>
</tr>
<tr>
<td>Light Brown</td>
<td>10%</td>
</tr>
<tr>
<td>Red</td>
<td>7%</td>
</tr>
<tr>
<td>Yellow, Gold</td>
<td>6%</td>
</tr>
</tbody>
</table>
c. Here are similar data for luxury cars made in Europe. Make a graph of these data.

d. What are the most important differences between choice of colors in Europe and North America?

<table>
<thead>
<tr>
<th>Color</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>30%</td>
</tr>
<tr>
<td>Silver</td>
<td>24%</td>
</tr>
<tr>
<td>Gray</td>
<td>19%</td>
</tr>
<tr>
<td>Blue</td>
<td>14%</td>
</tr>
<tr>
<td>Green</td>
<td>3%</td>
</tr>
<tr>
<td>White, Pearl</td>
<td>3%</td>
</tr>
</tbody>
</table>

7. Feeding at a carcass leads to competition among lions. Ecologists collected data on feeding contests in Serengeti National Park, Tanzania. In each contest, a lion feeding at a carcass is challenged by another lion seeking to take its place. Who wins these contests tells us something about lion society. The following table represents data on 396 contests between adult lion (male or female) and an opponent of a different class:

<table>
<thead>
<tr>
<th>Opponent</th>
<th>Adult Male</th>
<th>Adult Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contests</td>
<td>Male Wins</td>
</tr>
<tr>
<td>Adult Female</td>
<td>136</td>
<td>113</td>
</tr>
<tr>
<td>Sub-adult</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Yearling</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Cub</td>
<td>25</td>
<td>22</td>
</tr>
</tbody>
</table>

a. Make separate graphs for males and females that compare their success against different types of opponent.

b. Describe the most important differences between the behavior of female and male lions in feeding contests.

REPRESENTING QUANTITATIVE DATA

8. The Fathom dotplot below displays the widths (in centimeters) of the feet of 39 fourth-grade students from an elementary school in Georgia.

a. What is the foot width of a “typical” student in this group of 39 fourth-graders? Explain.
Here is another Fathom graph that displays the food widths of boys and girls separately.

b. Do fourth-grade boys have wider feet than fourth-grade girls? Give appropriate evidence from these side-by-side dotplots to support your answer.

9. Students in a high school statistics class were asked to report how many siblings they have. Here are the data

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

   a. Construct a dotplot to display these data.
   b. How many siblings does a “typical” student in the class have? Justify your answer.
   c. Are there are any really unusual values in the data set? Explain.

10. In a rural town in Oklahoma during the 1970’s, the following data was collected concerning the age at which the eldest child in a family to get his/her license. The sample consists of 31 people, 16 males (M) and 15 females (F).

M  16 16 17 16 18 17 17 16 16 27 16 17 16 17 16 16 16 16
F  17 18 19 20 18 19 20 18 18 17 16 18 19 17 18 17 18

   a. Calculate the following statistics for the male and female data separately.
      i. Mean
      ii. Median
      iii. Mode
      iv. Range
   b. Plot each set of data. Either use a stem-and-leaf plot or a dot plot.
   c. Looking at your plot, which data point(s) would be considered an outlier?
   d. Re-calculate part (a) above for the males, excluding the outlier if there is one.
   e. Summarize the effects of the outlier on the mean, median, mode, and the range.
   f. In statistics, often you will be required to interpret your data. In a paragraph, compare the two sets of data.
11. Suppose a set of data consists of 33 whole number observations. Its five number summary is \((\text{min}, Q_1, \text{median}, Q_3, \text{max})=(16, 20, 22, 30, 46)\). (This problem refers to box and whisker plots)
   a. What is the range of the data?
   b. How many observations are strictly less than 22?
   c. Is it possible that there is no observation equal to 22? (explain briefly)
   d. How many observations are strictly less than 20?
   e. Is it possible that there is no observation equal to 20? (explain briefly)
   f. Construct a modified box plot.

12. A study of computer-assisted learning examined the learning of “Blissymbols” by children. Blissymbols are pictographs (think of Egyptian hieroglyphs) that are sometimes used to help learning-impaired children communicate. The researcher designed two computer lessons that taught the same content using the same examples. One lesson required the children to interact with the material, while in the other the children controlled only the pace of the lesson. Call these two styles “Active” and “Passive.” Children were assigned at random to Active and Passive groups. After the lesson, the computer presented a quiz that asked the children to identify 56 Blissymbols. Here are the numbers of correct identifications by the 24 children in the Active group and the 24 children in the Passive group.

   a. Describe any similarities and differences you see in the dotplots. Be specific.
   b. Is there good evidence that active learning is superior to passive learning? Explain.

**DESIGNING AN EXPERIMENT**

13. Does studying while listening to music help or hinder learning? You have been asked to direct a study to investigate this issue. Would you recommend a survey, an experiment, or an observational study? Explain carefully why you chose the design you did, and why you did not choose the other two possible designs.

14. What percent of young adults regularly talk on a cell phone while driving? You have been asked to direct a study to investigate this issue. Would you recommend a survey, an experiment, or an observational study? Explain carefully why you chose the design you did, and why you did not choose the other two possible designs.

15. What percent of teenagers pray every day? You have been asked to direct a study to investigate this issue. Would you recommend a survey, an experiment, or an observational study? Explain carefully why you chose the design you did, and why you did not choose the other two possible designs.
16. A statistic is a number calculated from data. Quantitative data has many different statistics that can be calculated. Determine the given statistics from the data below on the number of homeruns Mark McGuire has hit in each season from 1982 – 2001.

<table>
<thead>
<tr>
<th>70</th>
<th>52</th>
<th>22</th>
<th>49</th>
<th>3</th>
<th>32</th>
<th>58</th>
<th>39</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>65</td>
<td>42</td>
<td>29</td>
<td>9</td>
<td>32</td>
<td>9</td>
<td>33</td>
</tr>
</tbody>
</table>

Using a TI-83 or TI-84, press STAT and Edit and enter these numbers in a list. Then go back to STAT, Calc, and 1-Var Stats to find the following (you can also do this without a calculator):

Mean (\(\bar{x}\)) ____________________
Standard Deviation (\(\sigma\)) ___________
MinX _______________________
Q1 _________________________
Median _____________________
Q3 _________________________
MaxX ______________________
Range ______________________
Q3 – Q1 _____________________

a. Which measure of central tendency best describes the data: mean, median, or mode? Explain.

b. Which measure of dispersion is best used to describe the spread of the data: range, interquartile range or standard deviation?

17. Calculator Problem:
Use the data below to do the following exercises:
12 18 33 15 66 43 26 42 22 4 17 54 44 52 38 27 25 44 18 25 42 63 38 36 44 19 58 30 14 46 21 34 54 23 44 55 60 39 62 33 28 53 35 55 44 21 55 50 33 9

a. Enter the data into a list (preferably L1).
b. Sort the list in ascending order and find the mode.
c. Find the mean of the data.
d. Find the median of the data.
e. Find the sum of the data.
f. Find the Standard Deviation stdDev() of the data.
REVIEW OF NECESSARY SKILLS:

NORMAL CURVE

Here is a formula that is used often in AP Statistics. It calculates a z-score (how many standard deviations away from the mean a certain piece of data lies).

\[ z = \frac{x - \mu}{\sigma} \]

18. If \( z = 2.5 \), \( x = 102 \) and \( \mu = 100 \), what is \( \sigma \)? Show your work. __________

19. If \( z = -3.35 \), \( x = 60 \), and \( \sigma = 4 \), what is \( \mu \)? Show your work.___________

LINEAR FUNCTIONS

20. The following data describes the number of persons per household in the United States in the census years between 1850 and 2000.

<table>
<thead>
<tr>
<th>Year</th>
<th># of persons in household</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td>5.55</td>
</tr>
<tr>
<td>1860</td>
<td>5.28</td>
</tr>
<tr>
<td>1870</td>
<td>5.09</td>
</tr>
<tr>
<td>1880</td>
<td>5.04</td>
</tr>
<tr>
<td>1890</td>
<td>4.93</td>
</tr>
<tr>
<td>1900</td>
<td>4.76</td>
</tr>
<tr>
<td>1910</td>
<td>4.54</td>
</tr>
<tr>
<td>1920</td>
<td>4.34</td>
</tr>
<tr>
<td>1930</td>
<td>4.11</td>
</tr>
<tr>
<td>1940</td>
<td>3.67</td>
</tr>
<tr>
<td>1950</td>
<td>3.37</td>
</tr>
<tr>
<td>1960</td>
<td>3.35</td>
</tr>
<tr>
<td>1970</td>
<td>3.14</td>
</tr>
<tr>
<td>1980</td>
<td>2.76</td>
</tr>
<tr>
<td>1990</td>
<td>2.63</td>
</tr>
<tr>
<td>2000</td>
<td>2.59</td>
</tr>
</tbody>
</table>

a. Construct a scatter plot of this data on graph paper. Put year on the x-axis and household size on the y-axis. Clearly label the axis, and give the graph a title.
b. Using a straight edge or a ruler, carefully draw a line of best fit through the data.
   i. Estimate the slope of your best-fit line. Include proper units for your slope.
   ii. Use your graph and best-fit line to estimate the number of persons in a household in 2010.
21. 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.

22. 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.

**SIMPLE PROBABILITY**

23. 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. 1/8   B. 2/9   C. 2/7   D. 3/8   E. 1/2

24. 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 five minutes.

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

A. 1/6   B. 1/3   C. 1/4   D. 1/2   E. 1

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

A. 1/8   B. 1/6   C. 1/4   D. 1/2   E. 1

27. 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 is 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: ___________, ___________, ___________, ___________
28. Write 7% as a decimal.

29. Write 3/10 as a percent.

30. What is 25% of 500?

31. What is 12% of 82?

32. What is the probability of rolling a number greater than a 2 when rolling a fair 6-sided dice?

33. What does 3<x<20 mean if you were to put in words?

34. \( G(x) = 3x - 5 \), find \( G(8) \).

35. Solve the following for \( z \): \( b = r(z) + g \)

36. Solve the following for \( z \): \( b = r(z) + z \)

37. Graph \( y = 4x + 3 \)

38. Graph \( y = 3 - 5x \)

39. Graph \( y = 22.4 + 6.8x \)

40. If the height of a human in inches is linearly related to their age in years, what would be the interpretation of a slope of 5.67?

41. Suppose that the scenario in #40 was represented by the equation \( A = 18.4 + 5.67h \). What is the value 18.4 represent in the real world?