Welcome to AHS Statistics!

I am looking forward to having you in class this fall. In order to be ready to hit the ground running in September, please utilize a free online Statistical tutorial to become familiar with descriptive statistics, types of data, and graphical displays.

In this class you will learn to describe and analyze sets of data and use that analysis to draw conclusions in context about the situation that gave the data. This course is not like any other math class you have ever taken! I would say that it is a combination of Math, English, and Science. Communication skills are essential, and there is much more reading and writing than what you are used to in a math class. It is a very rewarding course and a very important one, in my opinion, but can be quite difficult at times. This is considered a college-level class. The mathematics required for this course may not be as difficult as in other advanced math courses, but some of the concepts can be very confusing.

Preliminaries:

1. You must be competent with algebra concepts and skills.
2. You should be familiar with basic statistical plots such as box-and-whiskers plots, bar graphs, histograms, circle (pie graphs), and stem-and-leaf plots.
3. You must be willing to explain your answers. Just getting the correct number will not be sufficient for this course.
4. You must own your own calculator and bring it to class everyday. A TI-84 is an essential tool for this course, as it has many statistical features we will utilize. Please do not purchase a calculator unless it is a TI-84! Calculators can be rented from the library if you cannot afford to buy one.

The summer assignment is designed for you to review some algebra skills, refresh your knowledge or learn about basic descriptive statistics, graphs, and general information. Please write neatly. **This assignment is due the first day of class and will be graded.** It is recommended that you complete this assignment in August so the content is fresh in your mind when the school year starts. Transferring students are understandably allowed some leeway to catch up with their peers.

Please feel free to email me at pero@astoria.k12.or.us. I will check my email periodically throughout the summer.

Have a great summer and I'll see you in September.

Mr. Ero

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 information on the site you will be completing a vocabulary list (on preceding pages). Follow the steps below:
    - Go to www.stattrek.com
    - Click on "tutorial"
    - On the left side of the screen is a list of tutorials available. Click on the AP Statistics tutorial.
    - A list of general topics will appear on the left hand side of the screen. When you click on the general topic a list of subtopics will appear. You will read or watch the video on the following subtopics to complete the vocabulary list.

General Topic: The Basics	
Subtopics:	Variables
	Populations vs. Samples
	Mean and median
	Variability
	Position
General Topic: Charts and Graphs	
Subtopics:	Patterns in data
	Dotplots
	Histograms
	Stemplots
	Boxplots
	Cumulative plots
	Scatterplots
	Comparing data sets
General Topic: Categorical Data	
Subtopics:	One-way tables
	Two-way tables

2. **Practice Problems.** After reading all the material above and from your book, you should be able to complete the questions in the remaining pages of this packet. You may do so in the spaces provided.

Part I: Vocabulary List

Please define each of the following terms from the information on the stattrek website.

1. Categorical Variables

2. Quantitative Variables

3. Univariate Data

4. Bivariate Data

5. Median

6. Mean

7. Population

8. Sample

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. Dotplot

18. Scatterplot

19. Difference between a bar chart and histogram

20. Stemplots

21. Boxplots

22. Quartiles

23. Range

24. Interquartile Range

25. Parallel boxplots

26. Difference between a frequency table and relative frequency table

27. Difference between frequency and cumulative frequency

28. Parameter

29. Statistic

30. Marginal Distribution

31. Conditional Distribution

32. Segmented Bar Chart

33. What are the W's of data

Part 2: Practice Problems

CATEGORICAL OR QUANTITATIVE

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

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																

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

WHERE DO OLDER FOLKS LIVE?

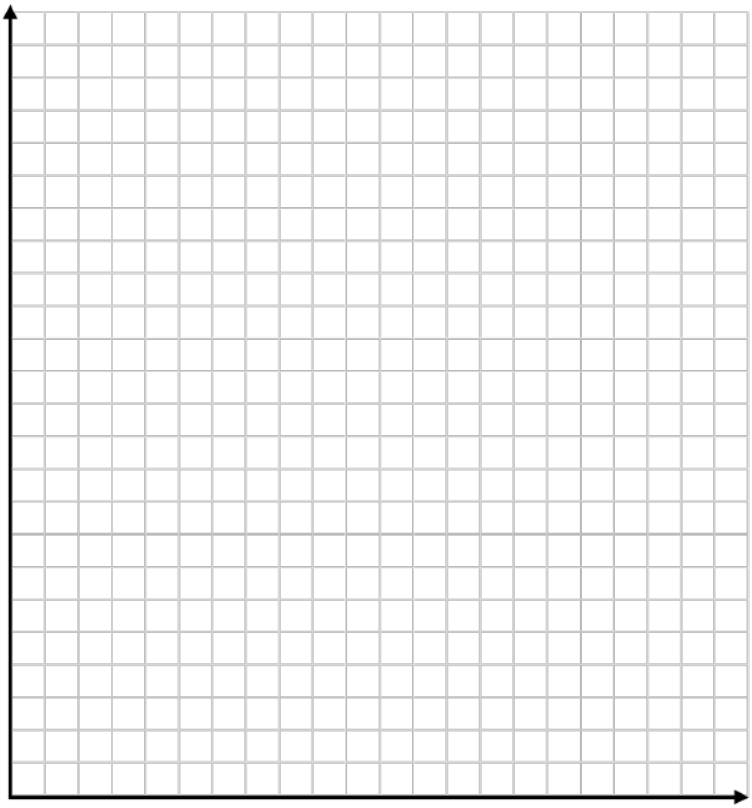
This table gives the percentage of residents aged 65 or older in each of the 50 states.

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.

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.

Bin Widths	Frequency
4 to < 6	1
6 to < 8	
8 to < 10	



SSHA SCORES

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

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

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

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.

New Grading Policy

A new grading policy has been proposed by the dean of the College of Education for all education majors. All faculty and students in the college were asked to give their opinions about the new policy. The results are given below.

	Favor	Neutral	Opposed	Row Total
Students	353	75	191	619
Faculty	11	5	18	34
Column Total	364	80	209	653

a. State the variables and if they are categorical or quantitative.

b. What percent of responses were from students favoring the policy change?

What percent of students favored the policy change?

What percent favoring the policy change were students?

c. What is the marginal distribution of the grading policy change?

d. What is the distribution of the grading policy among just students?

e. What is the distribution of the grading policy among just faculty?

f. Create a segmented bar graph of students and faculty and their view on the proposed grading policy change.

"Teen Automobile Crash Rates Are Higher When School Starts Earlier"

Science Daily (June 10, 2010) — Earlier school start times are associated with increased teenage car crash rates, according to a research abstract presented June 9, 2010, in San Antonio, Texas, at SLEEP 2010, the 24th annual meeting of the Associated Professional Sleep Societies LLC.

Results indicate that in 2008 the teen crash rate was about 41 percent higher in Virginia Beach, Va., where high school classes began at 7:20 a.m., than in adjacent Chesapeake, Va., where classes started more than an hour later at 8:40 a.m. There were 65.4 automobile crashes for every 1,000 teen drivers in Virginia Beach, and 46.2 crashes for every 1,000 teen drivers in Chesapeake.

"We were concerned that Virginia Beach teens might be sleep restricted due to their early rise times and that this could eventuate in an increased crash rate," said lead author Robert Vorona, MD, associate professor of internal medicine at Eastern Virginia Medical School in Norfolk, Va. "The study supported our hypothesis, but it is important to note that this is an association study and does not prove cause and effect."

The study involved data provided by the Virginia Department of Motor Vehicles. In Virginia Beach there were 12,916 drivers between 16 and 18 years of age in 2008, and these teen drivers were involved in 850 crashes. In Chesapeake there were 8,459 teen drivers and 394 automobile accidents. The researchers report that the two adjoining cities have similar demographics, including racial composition and per-capita income.

1) Answer the following questions regarding the above excerpt:

a) *Who* is being studied?

b) *What* about those individuals is being recorded / analyzed (i.e. what are the variables?)? Do you think the variables are categorical or quantitative in nature?

c) *When* was the data collected?

d) *Where* was the data collected (more accurately: what geographical area is associated with the data)?

e) *Why* do you think this data was collected and analyzed?

f) *How* was the data collected and analyzed? In other words, what methods were used?

g) Why do you think the authors of the study mentioned, "it is important to note that this is an association study and does not prove cause and effect?"

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 five minutes.

3. 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. ¼
   D. ½
   E. 1

4. 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

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: _____, _____, _____, _____