Welcome to AP Statistics

Required Materials:

- Pencils, colored pencils/pens, loose leaf paper, graph paper, 3 ring binder & dividers
- Graphing Calculator (TI-83/84 plus family)

*It is **highly recommended** to purchase a **TI-84 plus** (or equivalent) calculator. If the student has a graphing calculator (i.e. TI-83) it may be used as well. Classroom calculators (if provided) must be returned at the end of class each day.

Prerequisites: Successfully completed Algebra I, Algebra II, and Geometry with a "B" or higher and passed the Algebra I HSA/PARCC.

Skills Required:

- Write, evaluate, and interpret algebraic equations.
- Choose, use, and substitute values in for statistical formulas.
- Use TI-84 plus (or TI-84 family) graphing calculator.
- Synthesize information from text, and write in full, clear, concise sentences.
- Interpret information from tables and graphs.

Student Responsibilities: Reminder… This is a COLLEGE level course!

- Read the textbook and take notes daily (before, during, and after class).
- It is YOUR RESPONSIBILITY to make up all work missed for each **excused absence** from class and turn it in no later than one class day for each day absent.
- Study without the direction of the teacher.

Welcome Back Assignment

Due by: September 13th 2019

Your welcome back assignment consists of two parts. Part I is a review of Algebra I Data Analysis, and part II is a written assignment. The purpose of part I is to determine your prerequisite knowledge and current level of understanding of some key points in statistics and data analysis that we will investigating further during the course. Feel free to use outside resources including the internet to help you answer those questions. If at any time you are struggling with the assignment, feel free to contact me via email:

Mrs. Rivoli – achiodi@bcps.org

Period:______ **Name:____________________________________**

AP Statistics Summer Assignment Grading Rubric

Part I: (selected problems)	**Score**
Algebra & Data Analysis	___ / 8
Boxplots	___ / 10
Probability	___ / 8
Part II:	
Font/Title/Headings, Mechanics, Word Length, & Citation	___ / 4
Introduction	___ / 5
Article Review	___ / 10
Research & Apply	___ / 5

Total Score: ___ / 50

Additional Comments:

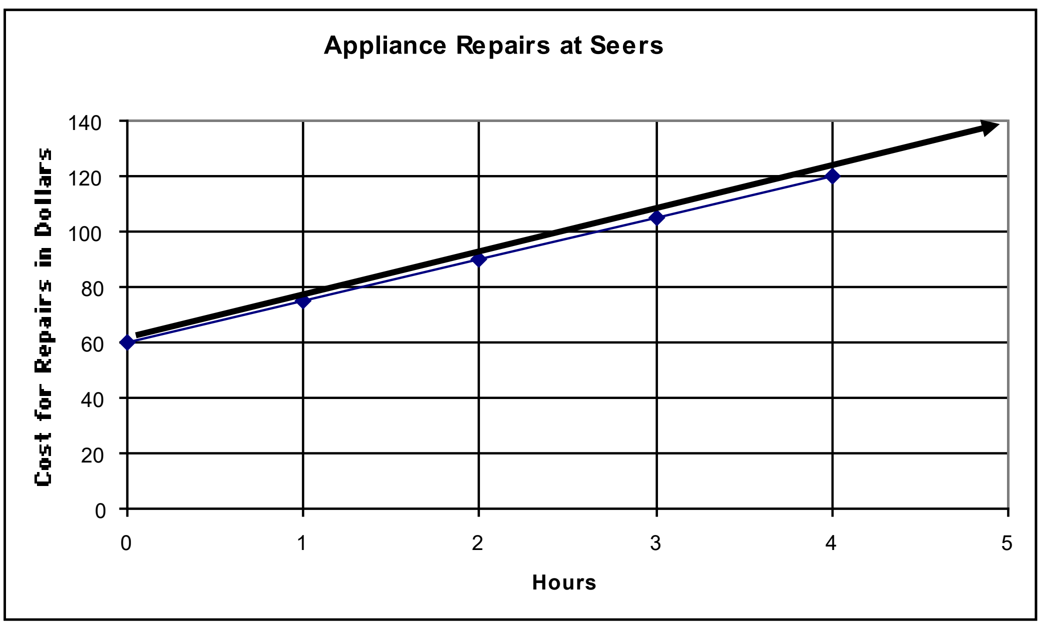
Date:________________ Period:_______ Name:____

Part I
Algebra HSA & Data Analysis Skills Review

Linear Regression

1. Mr. Green rented a truck full of hay for a hayride. He used the equation $p = 3.5x - 85$ to determine profit for selling x tickets. What is the slope of the equation? What does the slope represent in the context of the problem?

2. Seers department store uses a graph to show the cost of repairing appliances. Seers charges $15 per hour for each repair.



What does the y-intercept represent in the context of this problem?

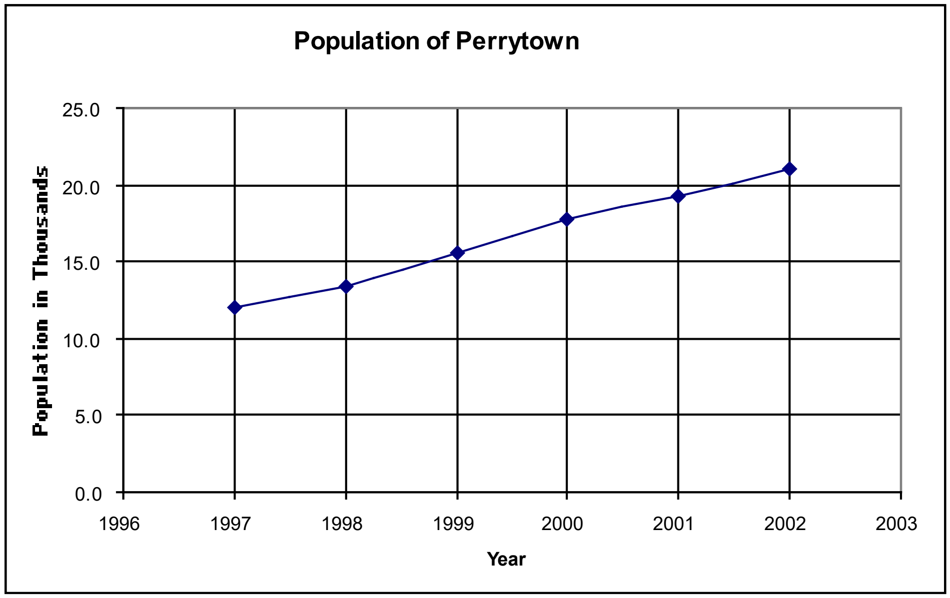
A. the cost per hour
B. the minimum charge
C. the number of hours
D. the number of appliances

3. Given the following table, what is the value of y when x = –2?

x	–1	0	1
y	$-\frac{9}{2}$	–4	$-\frac{7}{2}$

4. The graph below shows the population of Perrytown (in thousands) from 1997 to 2002.

Year	Population
1997	12.0
1998	13.4
1999	15.6
2000	17.8
2001	19.3
2002	21.0



- Calculate the average rate of change from 1997 to 2002.

- Assuming the rate of change remains constant, predict the town's population in the year 2003. Use mathematics to explain how you determined your answer. Use words, symbols, or both in your explanation.

- Predict the population in the year 2005.

5. The table below shows the cost of computers.

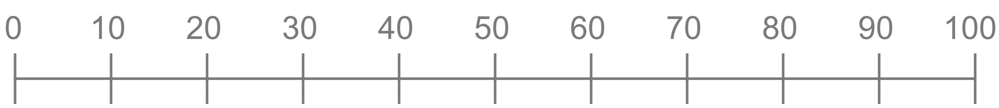
Year	1998	1999	2000	2001	2002
Cost	$2,345	$2,008	$1,875	$1,625	$1,239

- Use your graphing calculator to find the equation for a line of best fit. Write it below:

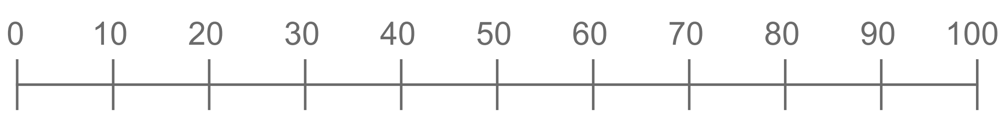
Box Plots

- Create a box plot for each set. Also compute the mean and median values for each set.

1. 12, 32, 56, 81, 83, 86, 86, 91, 93, 95, 96



2. 32, 23, 56, 84, 78, 81, 93, 46, 56, 74, 100



3. Students sold gift wrap for a club fundraiser. The following table represents the amount of gift wrap each student sold.

Student	Amount Sold	Student	Amount Sold
Mary	$123	Allison	$85
Tom	$236	Kelly	$321
Frank	$282	Rachel	$211
Dwayne	$1,426	Susan	$73
Marge	$325	John	$421
Paul	$96	Matthew	$386

Construct a box plot of the data to help answer the following.

a. Which students' sales are above average?

____

____

b. Which students' sales are below average?

____

____

c. Which students have average sales?

____

____

d. Is the mean or the median a better representation of the average? Use mathematics to justify your answer.

____

____

____

Probability

1. The administration needs to pick an eleventh or twelfth grade student from the SAT review classes to attend a free SAT seminar. There are 65 eleventh and 85 twelfth grade students in these classes.

   a. What is the probability that an eleventh grade student will be selected in a random drawing?

   b. What is the probability that a twelfth grade student will be selected in a random drawing?

2. Major Motors made a new car model. You may select from 8 different colors of paint, 4 cylinder engines or 6 cylinder engines, and 3 different shades for the interior of the car. How many different varieties of the new model car are there?

3. A survey shows $\frac{1}{4}$ of the drivers prefer SUV's. What device could be used to simulate the probability that a driver prefers an SUV?

   A. A Six-sided number cube
   B. A Spinner with 3 equal sections
   C. A fair coin
   D. A standard deck of 52 playing cards

4. A poll shows $\frac{1}{6}$ of the voters are Independents. What device could be used to simulate the probability that a voter is an independent?

   A. A Six-sided number cube
   B. Spinner with 3 equal sections
   C. A fair coin
   D. A standard deck of 52 playing cards

5. The table below shows approximate earnings for parents of children in a small preschool.

Earnings (in thousands)	40	50	60	70	80	90	100
Number of parents	2	10	23	13	8	2	2

What is the probability that a parent selected from this group would make $70,000 or more?

6. The Department of Education wants to test the mathematical ability of students at a high school. According to simple random sampling techniques, which of the following is the best way to randomly sample 80 students? Use mathematics to justify your answer.

A. Select the first 80 students who come to the cafeteria in the morning.

B. Select 80 students from the band

C. Ask algebra teachers to randomly select 80 students

D. Give every student in the school a number and use a calculator to randomly generate 80 numbers.

7. The stem-and–leaf plot below shows the number of cars sold by Arrow Motors each month they have been in business. What is the probability that Arrow Motors sold more than 72 cars during a randomly selected month?

Number of Cars Sold by Arrows Motors

5	1					
6	2	2	5	5		
7	1	1	3	5		
8	3	4	6	6		
9	1	2	8	9	9	

Key 5|1 = 51 cars

Part II

"The Untruth about Statistics"

Introduction

1. Why are you taking AP Statistics? Is statistics an important tool in the field of study you are considering? Have you had any previous connection with statistics? Can you give an example from the "real world" where statistics is or has been important to you personally? This is how I will get to know you, as a student.

Brief Article Review

2. Read the article *The (un)Truth about Statistics*.
3. Create a response for the article. Your response can be in the form of a reflection (I agree or disagree – answer why), or a summary that includes a thoughtful analysis of the article.
   - The main ideas and points of the article need to be included.
   - Each response needs to be at a minimum 500 words in length. Grammar and spelling need to be considered when writing responses.
   - I am looking for responses that are well thought out. Written responses should not be just a regurgitation of what was presented in the article. Have fun with this and be creative.

Research & Apply

4. Find an article, study, commercial, advertisement, etc. that uses statistics and identify whether or not you believe the statistical claims made are legit. Use the main ideas from *The (un)Truth about Statistics* and your current knowledge of statistics to help support your beliefs.
   - You must site your source (website, television network, product name, study, magazine, etc), the location of your source, and the date you found your source. This can be informal.
   - Responses must be a minimum of 150 words in length.

Requirements
In order to demonstrate knowledge facility with certain features of your word processor, after you finish writing your paper, I want you to:

1. Make the first line the title of your assignment **The Untruth about Statistics** (bold) and make the second line (of the title) your name. Center both lines and make the title 14 pt. Times New Roman font.
2. The body of your response should be to 12 point Times New Roman font. If this is not available on your computer, use a simple newspaper-type font like the one you are reading.
3. Insert the following paragraph headings in **bold**, in order: **Introduction, Article Review,** and **Research & Apply** like the paragraph headings I have used on this sheet.
4. Before you print the final copy, make sure you proofread your paper on the screen, and then check it for spelling and syntax.

The (un)Truth About Statistics

Have you ever heard the expression, "Four out of five doctors recommend..."? Or "... 42% more relief from heartburn"? Or "... better highway mileage than any other sub-compact hatchback sedan costing under $10,000 made in America"?

Perhaps you suspected that these claims were not completely true. It is wise to be suspicious, because **statistics** (and numbers in general) can be manufactured to make any idea sound convincing. When used properly, statistics is a powerful tool for uncovering truth; when used improperly, it can be manipulated to prove almost anything.

Try, try again

There are lots of ways to misuse statistics. One way is perseverance: if at first you don't succeed (i.e., get the result you wanted), try, try again. Suppose you want to claim in a TV commercial that 4 out of 5 dentists recommend your toothpaste. You ask 5 dentists, but only 1 of them recommends your brand. So, forget you ever asked them! Ask another 5 dentists! This time, 2 of them recommend your brand. Forget them! Ask another 5! Keep trying until, by random fluctuation, you get lucky and 4 out of 5 recommend your brand. Then, show your TV commercial. Whatever you do, *do not* talk about the 13,925 dentists you had to survey before you got lucky, and don't mention that only 8% of *them* recommended your brand.

Sometimes this sort of thing happens even to honest people. If the results do not match our theory, it is too easy to think of a "good" reason to believe that data we do not like are not valid, so we have to do the experiment again. This happens far too often in scientific research, even today. Despite people's best intentions to be fair, there is just too much temptation to rationalize away the "bad" data. However, you rarely see any scientists rationalize away the "good" data, the data which support their theories!

Here we have the first lesson of honest statistics: you cannot ignore the data that do not fit your theory. Sometimes you have good reason to believe some piece of data should be excluded because it is just a mistake. But in your scientific report, you have to say so, and state exactly why it has been omitted. You can exclude data if you have good reason, but you cannot ignore them, or fail to report them.

How many?

A nursing home recently tried new procedures designed to reduce the number of accidental injuries to patients. They were pleased to announce that in the first four months of the year, patient accidents were down a whopping 60% compared to last year. Can't argue with that!

Or can you? How many are we talking about here? If last year there were 50 accidents, and this year only 20, then they are down 60%, and there is no doubt that this result is statistically significant. The chance of the result happening by random fluctuation ("by accident") is less than 1 in 10,000.

But suppose there were 5 accidents last year, and only 2 this year. Yes, they are down 60%. But no, this result is *not* significant. The chances are better than 1 in 4 that this could happen by random fluctuation.

We have already seen that as we acquire more data, our results become more precise. They also become more *reliable*. Sometimes, an early result is based on so little data that it has no real significance. Do not put too much faith in statistical results (not even a whopping 60%) until you know how much data went into them.

Survey says!

Suppose two politicians are debating a school funding bill. They both try to show that the public is on their side by conducting a survey. Politician **A** wants to show that people favor the bill, so his survey asks, "Should we invest more in our children's future by passing the school funding bill?" Lo and behold, people *do* want to invest in their children's future, so most people say yes, and politician **A** announces that the vast majority favor his bill.

Politician **B** wants the bill to fail, so his survey asks, "Should we raise taxes to fund more and bigger government bureaucracy by passing the school funding bill?" Not surprisingly, people do not want higher taxes and more bureaucracy, so they mostly say no, and politician **B** claims that the vast majority oppose the bill.

This may seem like an exaggerated example, but it is not. This actually happens! Almost every political survey is deliberately designed to produce a specific response. The questions are usually phrased to make the desired response sound good, while making the undesired response sound very bad. By doing so, the questions bias the subject's opinion about the topic of the survey. Not surprisingly, whoever paid for the survey usually gets the response they want. Politicians are not the only ones who do this. Advertising surveys are carefully designed to make the company product look good while making the competition look bad.

Even if you are trying very hard to be fair, it is actually quite difficult to phrase the question in a way that does not influence anyone's response. There are other ways surveys can go wrong, too; designing an accurate survey is a very difficult task, requiring much expertise. There are some organizations that do it well; for example, the Gallup organization specializes in conducting fair, scientifically reliable surveys. Still, it is an unfortunate fact that *most surveys just cannot be trusted* (especially political and advertising surveys).

What are you trying to prove?

It happens regularly that a government agency or private commission launches a major study of an important social issue. Too often they begin by announcing that they are going to prove some theory, which has important consequences for social policy. You can bet big money that they *will* find proof. After all, they have already made up their minds!

Any study which begins by assuming the correct answer, then looks for proof, will fail to give serious consideration to the possibility that the assumed "correct answer" is *not* correct. Any scientist who has already decided before the experiment that one result is "right" and another is "wrong" is no scientist at all.

It is very hard to avoid all bias when taking data. That is why we work very hard to make our experiments **double blind**: we arrange that neither the scientists taking data, nor their subjects, know how the data will affect the outcome. For example, suppose we want to study the effectiveness of a new headache pill. We give half our subjects the new medication, while the other half are given an inert sugar pill. We have to be sure that the subjects *do not know* which one they are getting. We also have to be sure that the scientists taking the data also do not know (at least until all the data are in). Otherwise, there is far too much temptation to "nudge" the data the way we want them to go.

Accidents happen

We have said that the standard of "unlikeliness" in statistics is 0.05, or 5%, or a 5% **false-alarm probability**. This means that if we do a scientific experiment, and get a result that's only 5% likely to happen by accident, we have evidence that it is not an accident. We can write our results in a scientific paper, and every statistician will agree that our evidence is significant.

So we have evidence, but we do not yet have *proof*. After all, there *is* a 5% chance that it *did* happen by accident. Accidents do happen! In fact, an accident that is only 5% likely will happen about 5% of the time. After all, with a 5% false-alarm probability, we will get some false alarms.

Suppose a university employs 100 scientists, and each one does a different scientific experiment. From probability theory, we *expect* 5% of them to get a result that's only 5% likely, *by accident*! So *just by accident*, about 5 of the 100 scientists will get evidence that they can call "statistically significant" and publish in a scientific paper.

And they *do* have evidence, strong enough that their claim deserves further study. But they do not have proof. That is one of the reasons scientific experiments have to be *repeated*. If you get a "significant result" once, you have evidence. If two people get the same result, there is very strong evidence. If a dozen people do the same experiment, and they all get a significant result, then we can start to believe it.

Every year, scientists do hundreds of thousands of experiments. If they use a 5% false-alarm probability (and most of them do), we can *expect* 5% of the results to be false alarms. Five percent of 100,000 experiments is 5,000 false alarms! That means 5,000 results that seem to be significant, but really happened only by accident. Some of them will be published in important scientific journals. And they should be published: they are all *possibilities*, and deserve further study. But for most of them, we should not be convinced until the results are repeated.

Conclusion

We have seen that if you want to deceive people, statistics makes it easy. In fact, even if you want to be honest, there are so many things that can go wrong in an experiment or a survey that we must carefully guard against bias. Even if we succeed, and get an unbiased result which is "statistically significant," it still might have happened just by accident. So the experiment has to be repeated, many times, and each time requires the same care in guarding against any bias which could affect the results.

That is a lot of work! Still, the payoff makes it well worth it. Not doing so gives us half-baked theories which sound good but really are not, supported by biased data and invalid statistics. This is worse than ignorance! But if we invest the effort to do science well, we reap the reward of knowledge that we can trust, and often can put to very good use.

March 12^{th}, 2006 by Zunairah

http://www.chowrangi.com/the-untruth-about-statistics.html