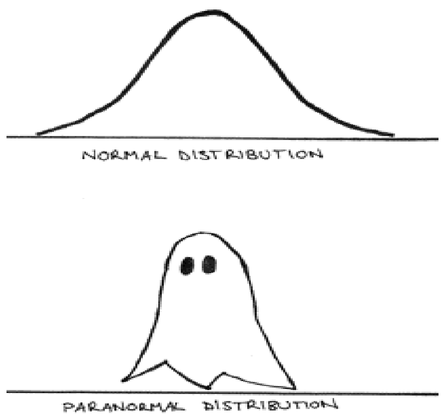
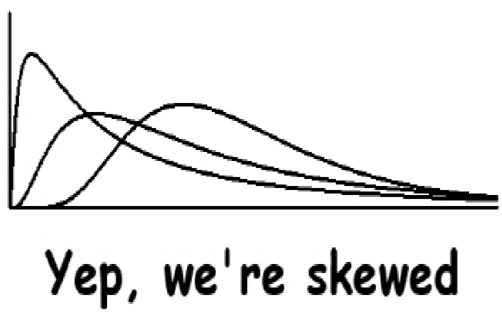
Name: ____

<u>AP Statistics Summer Project</u> - Ms. Klimczuk





<u>Instructions</u>:

- This project is due **<u>WEDNESDAY, AUGUST 15th at 1pm</u>**. You may pass it in early if you prefer. You will need to come to school to pass it in to my mailbox in the main office.

- If you need help with your summer project for any reason, please e-mail me. You may not get an immediate response, but I'll check my mail periodically and respond to questions – **<u>DON'T TURN TO A FRIEND FOR HELP, YOU'RE SUPPOSED TO WORK ALONE, SO ASK ME INSTEAD</u>**!!

  My Email: cklimczuk@oxps.org

  I am looking for EFFORT, not perfection.

- I expect you to research terms and questions with which you're not familiar. Words in **<u>BOLD</u>** are key terms – it's worthwhile to look these up to ensure you know what they mean and how they're used in AP Statistics. This is where you will use your text book to help with your project.

- Several sections of this assignment ask for typed up responses. Use separate paper for these questions, type your answers, label/number each response carefully, use proper grammar and correct spelling.

- This will be counted as a TEST grade!

*****<u>Late submissions will be accepted on the first day of class for half credit at most.</u>*****

<u>Section I</u>: Reflection

Type a one paragraph reflection about why you are taking AP Statistics next year. Please be honest and thoughtful about your response.

<u>Section II</u>: **Watch and summarize the "AP Stats Guy" summer videos.** Go to www.apstatsguy.com [Get Yo Five!] and watch Summer Videos one through Four. Do **<u>NOT</u>** complete the summer assignment posted there, just watch the videos.

Type a few (3-5) sentences summarizing what you learned from **<u>EACH</u>** of the summer videos you watched at www.apstatsguy.com. It should be abundantly clear from your summaries that you did, in fact, watch the videos. Label each response "Summer Video One," Summer Video Two, etc.

The rest of this project explores the visual display and analysis of data to determine students' ability, with the aid of a graphing calculator, to:

- Enter and manipulate data;
- Complete basic numerical computations (5-number summaries, mean and standard deviation);
- Create properly formatted visual displays of data (tables, bar graphs, histograms, box plots, stem-and-leaf plots);
- Analyze and interpret numerical summaries and visual displays of data.

Calculator Help for the TI-84 Series:

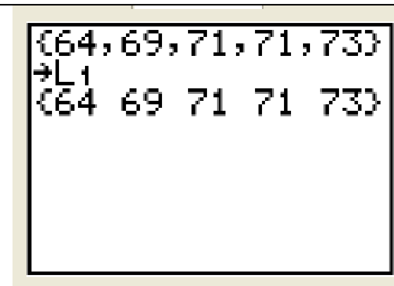
Entering data into a list

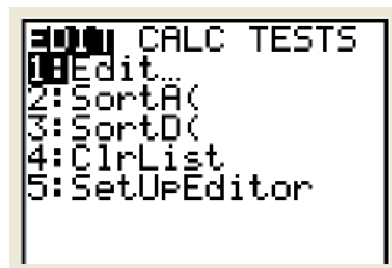
There are two ways to enter data into a list.

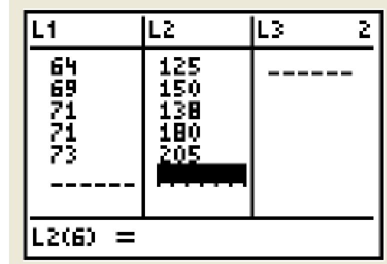
1. Enter the data directly from the home screen. To enter data this way, press [2nd], <{>, the value(s) to be stored separated by commas, <}>, [Sto>], list name, [Enter]
2. Enter the data using the Stat editor. To enter data this way, press [Stat], {Edit}, the appropriate list, the value(s) to be stored, [V] (down arrow) or [Enter] after each entry. Press [2nd], <Quit> when complete.

Examples:

- To enter the values 64, 69, 71, 71, and 73 into list 1, press: [2nd], <{>, 64 [,] 69 [,] 71 [,] 71 [,] 73 <}>, [Sto>], [2nd], <L1>, [Enter]
- To enter the values 125, 150, 138, 180, and 205 into list 2: press [Stat], {Edit}, {L2}, 125 [V] 150 [V] 138 [V] 180 [V] 205 [V] [2nd], <Quit>







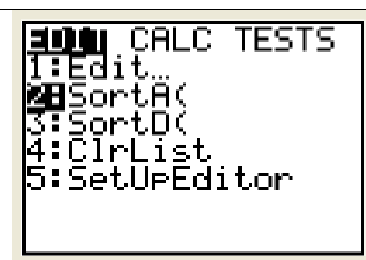
Other

Sorting lists

Lists may be sorted in either ascending or descending order.

1. To sort a list in ascending order, press: [Stat], {Edit}, {SortA(}, list name, [)], [Enter].
2. To sort a list in descending order, press: [Stat], {edit}, {SortD(}, list name, [)], [Enter].

Example: to sort L_1 in ascending order, press: [Stat], {Edit}, {SortA(}, [2nd], <L_1>, [)], [Enter].



1 Variable Statistics

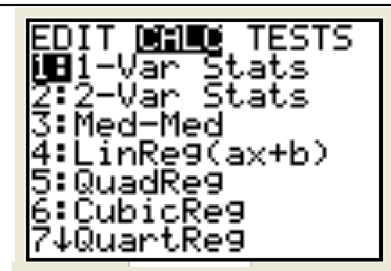
The 1-variable statistics function analyzes data with one variable. This function may be executed two ways.

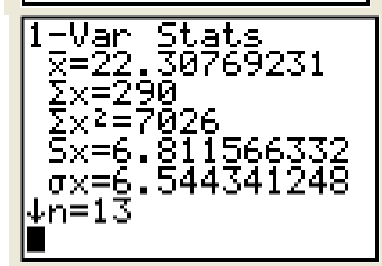
1. From a single list. This procedure is used if all of the data values are entered into a single list, with values being repeated if the value occurs more than once. Press: [Stat], {Calc}, {1-Var Stats}, location of data, [Enter].
2. From two lists. This procedure is used if one list contains the values and a second list contains the frequency of occurrence for each value. Press: [Stat], {Calc}, {1-Var Stats}, location of values, [,], location of frequencies, [Enter].

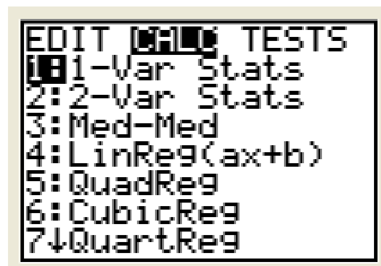
Examples:

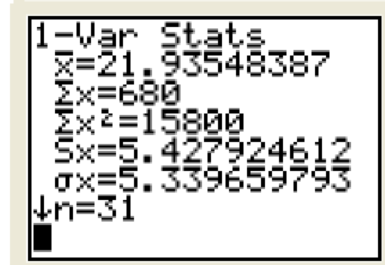
- To determine the 1 variable statistics for the data in L_3, press: [Stat], {Calc}, {1-Var Stats}, [2^{nd}], <L_3>, [Enter].

- To determine the 1 variable statistics for the data in L_3, with L_5 being the frequency of each occurrence, press: [Stat], {Calc}, {1-Var Stats}, [2^{nd}], <L_3>, [,], [2^{nd}, <L_5>, [Enter].









Section III: Numerical Summaries and Visual Displays (round all answers to the tenths place).

Data:
Scores (in percentages) for the first test of the year for two AP Statistics classes are provided below. The class "First" meets first period of the day, every day of the school year. The second class "Last" meets last period of the day, every day of the school year. There are 19 students in the First class and 30 students in the Last class.

First Scores: 93, 47, 87, 87, 55, 61, 85, 81, 81, 74, 86, 84, 84, 84, 76, 76, 78, 82, and 82 (***sum***: 1,483).

Last Scores: 94, 93, 90, 52, 55, 56, 84, 83, 80, 80, 70, 58, 64, 68, 68, 70, 82, 82, 79, 78, 78, 77, 76, 76, 76, 75, 74, 74, 72, and 72 (***sum***: 2,236).

The teacher would like to conduct an analysis comparing the grades of the classes to determine if there is a difference in skill level between the students in the classes.

Enter the data sets into two columns in the list function of your graphing calculator. Order the data from highest to lowest (run 1-var statistics and check the ***sum*** of each column to ensure that you entered your data correctly; the sum of each data set is given).

1. Use your calculator to find the **mean** and **standard deviation** for each class.

   First Mean: ______________ Last Mean: ______________

   First Standard Deviation: _____________ Last Standard Deviation: _____________

2. Which standard deviation did you record above, s or σ? Explain why: ________________________

   ____

   ____

   ____

3. Create a **5-number summary** for each class, round to 1 decimal place & record your results here:

	First	**Last**
Max		
Q3		
Median		
Q1		
Min		

4. Identify any **<u>outlier</u>**(s) in these data sets. Use **Tukey's Rule for Outliers** to support your answer.

5. Create a **<u>box plot</u>** (clearly showing any outliers) for each class. Sketch the box plots **<u>side-by-side</u>** (using the same scale).

6. Explain in a sentence or two the usefulness of a box plot when analyzing data.

7. Create **back-to-back stem-and-leaf plots** for the data sets.

8. Explain in a sentence or two the particular usefulness of a stem-and-leaf plot in analyzing data.

9. Fill in the **frequency table** of As, Bs, Cs, Ds and Fs for each class.

Grades	**FIRST**	**LAST**	**Totals**
A (90 +)			
B (80-89)			
C (70-79)			
D (60-69)			
F (< 60)			
Totals			

10. Using the table, calculate the following **probabilities**:

A. Probability of getting an A

B. Probability of getting an A if you're in the First period class

C. Probability of getting an A if you're in the Last period class

D. Probability that a selected student who has an A is from the Last period class

11. Create a **histogram** for each class.

12. Explain in a sentence or two the particular usefulness of histograms when analyzing data.

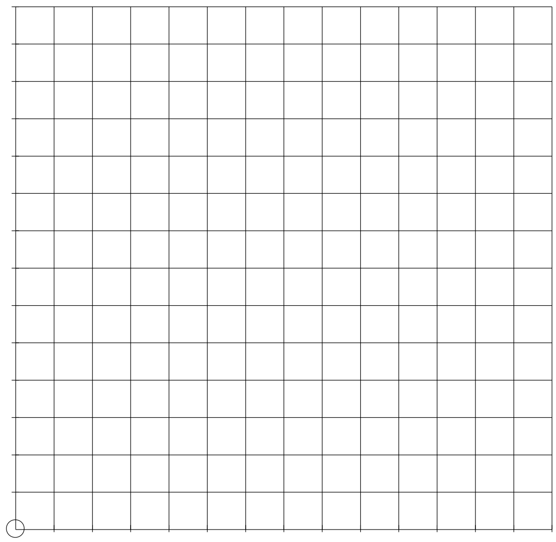
13. **Summary Questions: Using the data, tables, summary statistics from the previous pages, answer the following questions. Your answers should be 3-5 sentences each and typed.**

    A. Describe the **distribution** (**shape**, **center**, **spread** and **unusual features**) for each data set.
    B. Discuss your numerical findings in general, comparing the data of these two classes. What conclusions can you make?
    C. Should the AP Statistics teacher conclude that there is a difference in the level of abilities between the students in the two classes? Support your conclusion using your data.
    D. Are there factors besides student ability that might be affecting this data? Using your experiences as a student, identify some possible factors and support your argument.

The table below shows the number of basketball jump shots made in 25 attempts from a variety of distances.

X – Distance from basket (ft)	25	15	10	20	15	5	18	5	25
Y – Jump shots made	6	10	15	7	9	18	10	20	5

14. Label the axes and make a **scatter plot** of the data.



15. Does your graph show a positive, negative, or no **association**? ______________________________

16. Estimate the **correlation coefficient** to the nearest tenth. $r \approx$______________________________

17. How is correlation different from association?

18. Use a ruler to draw a **line of best fit** for your data.

A. Write the equation of the line here: ____

B. Estimate the number of jump shots made from a distance of 12 feet. __________________

C. Estimate the number of jump shots made from a distance of 50 feet. __________________

D. As a percentage, how confident are you about your answer to part C? Support your answer in a couple of sentences.

<u>Section IV</u>: Articles and Interpretation

Read the first article enclosed with this packet ("Research Basics: Interpreting Change").

Then read the second article ("Overstating Aspirin's Role in Breast Cancer Prevention") and answer the following questions - ***<u>typed, on separate paper and in complete sentences</u>***, please number your responses.

1. What was the story that the newspapers wrote about aspirin after the research was published by the Journal of the American Medical Association?

2. What other information should have been reported so that people could make decisions for themselves about the use of aspirin to prevent breast cancer?

3. How was the data collected to perform this study?

4. What type of study was performed?

5. Can this type of study be used to prove that aspirin prevents breast cancer?

6. What type of study must be done in order to ***prove*** something?

7. What is the difference between ***cause*** and ***association***?

8. You may have heard the statement "you can prove anything with statistics". Using what you have learned reading this article, explain what you think is meant by this statement.

The Washington Post

Research Basics: Interpreting Change

Tuesday, May 10, 2005

How Big Is the Difference?

Many medical studies end up concluding that two groups have different health outcomes -- death rates, heart attack rates, cholesterol levels and so forth. This difference is typically expressed as a *relative change* , as in the statement: "The treatment group had 50 percent fewer cases of eye cancer than the control group." The problem with this comparison is that it provides no information about how common eye cancer is in either group.

Thinking about relative changes in risk is like deciding when to use a coupon at a store. Imagine you have a coupon that says "50 percent off any one purchase." You go to the store to buy a pack of gum for 50 cents and a large Thanksgiving turkey for $35. Will you use the coupon for the gum or the turkey? Most people would use it for the turkey.

Why? Because paring half the price off $35 reaps a bigger savings --$17.50 --than cutting half off 50 cents -- or $0.25.

The analogy in health is that "50 percent fewer cases" is a very different number when applied to eye cancer -- a rare problem accounting for about 2,000 new cases in the U.S. each year -- than when applied to heart attacks -- a common problem accounting for about 800,000 new cases annually.

To really understand how big a difference is, you need to find out the *starting* and *ending points* -- sometimes called " *absolute risks* ." In the coupon example, the start and end points are the regular and the sales price. In a study about medical treatment, the start and end points are the chances of something happening in the untreated and treated groups.

Presenting the starting and ending point requires a few more words than presenting relative changes. For example, "In a year, two of 100,000 untreated people developed eye cancer; in contrast, one of 100,000 treated people developed eye cancer." For the price of a few more words you gain perspective: The chance of developing eye cancer is small.

Cause or Association?

Many important insights into human health come from *observational studies* -- studies in which the researcher simply records what happens to people in different situations, without intervening. Such studies first linked cigarette smoking to lung cancer and high cholesterol to heart disease. But not all observed associations represent cause and effect. And problems can occur when this key point is overlooked.

An example may help make the distinction clear. A man thought his rooster made the sun rise. Why? Because each morning when he woke up while it was still dark, he would hear his rooster crow as the sun rose. He confused association with causation until the day his rooster died, when the sun rose without any help.

A more serious example involves the long-held belief that most women should take estrogen after menopause. That idea, only recently discredited, also came from observational studies. The observation -- shown in more than 40 studies involving hundreds of thousands women -- was that women who took estrogen supplements also had less heart disease. But it turned out that estrogen was not the reason why this was the case. Instead, women taking estrogen tended to be healthier and wealthier. Their health and wealth -- not their estrogen supplements -- were responsible for the lower risk of heart disease.

The only way to reliably distinguish a cause from an association is to conduct a true experiment -- a *randomized trial* . In this type of study, patients are assigned randomly --that is, by chance--to receive a therapy or not receive it. This study design is the best way to construct two groups that are similar in every way except one -- whether they get the therapy being studied. That means any differences observed afterward must be caused by the therapy. In the case of estrogen and heart disease, such a study showed that the long-held beliefs were wrong.

Unfortunately, it is not always possible to do a randomized trial. For example, it is extremely unlikely that we could get people to agree to be randomly assigned to either eating only fast food or only organic food every day for a year (and that they would actually adhere to the diet if they did agree to be randomized). In such cases, scientists have to rely on observational studies. But when new tests or treatments are proposed, randomized trials ought to be conducted prior to their widespread use. Doctors prescribed estrogen to millions of women for many years until the randomized trial showed that intuition and dozens of observational studies were wrong.

The Washington Post

A May 10 Health section story about a study exploring aspirin use and breast cancer prevention incorrectly labeled hormone receptor positive cancers the most dangerous kind. That description applies to hormone receptor negative breast cancers.

Overstating Aspirin's Role in Breast Cancer Prevention
How Medical Research Was Misinterpreted to Suggest Scientists Know More Than They Do

By Lisa M. Schwartz, Steven Woloshin and H. Gilbert Welch
Special to The Washington Post
Tuesday, May 10, 2005

Medical research often becomes news. But sometimes the news is made to appear more definitive and dramatic than the research warrants. This series dissects health news to highlight some common study interpretation problems we see as physician researchers and show how the research community, medical journals and the media can do better.

Preventing breast cancer is arguably one of the most important priorities for women's health. So when the Journal of the American Medical Association published research a year ago suggesting that aspirin might lower breast cancer risk, it was understandably big news. The story received extensive coverage in top U.S. newspapers, including The Washington Post, the Wall Street Journal, the New York Times and USA Today, and the major television networks. The headlines were compelling: "Aspirin May Avert Breast Cancer" (The Post), "Aspirin Is Seen as Preventing Breast Tumors" (the Times).

In each story, the media highlighted the change in risk associated with aspirin -- noting prominently something to the effect that aspirin users had a "20 percent lower risk" compared with nonusers. The implied message in many of the stories was that women should consider taking aspirin to avoid breast cancer.

But the media message probably misled readers about both the size and certainty of the benefit of aspirin in preventing breast cancer. That's because the reporting left key questions unanswered:

· Just how big is the potential benefit of aspirin?

· Is it big enough to outweigh the known harms?

· Does aspirin really prevent breast cancer, or is there some other difference between women who take aspirin regularly and those who don't that could account for the difference in cancer rates?

This article offers a look at how the message got distorted, what the findings really signify--and some broader lessons about interpreting medical research.

How Big a Benefit?

Just how big is the potential benefit of aspirin?

The 20 percent reduction in risk certainly sounds impressive. But to really understand what this statistic means, you need to ask, "20 percent lower than what?" In other words, you need to know the chance of breast cancer for people who do not use aspirin. Unfortunately, this information did not appear in any of the media reports. While it might be tempting to fault journalists for sloppy, incomplete reporting, it is hard to blame them when the information was missing from the journal article itself.

In the study, Columbia University researchers asked approximately 3,000 women with and without breast cancer about their use of aspirin in the past. The typical woman in this study was between the ages of 55 and 64. According to the National Cancer Institute, about 20 out of 1,000 women in this age group will develop breast cancer in the next five years. Therefore, the "20 percent lower chance" would translate into a change in risk from 20 per 1,000 women to 16 per 1,000 -- or four fewer breast cancers per 1,000 women over five years.

For people who prefer to look at percentages, this translates as meaning that 2 percent develop breast cancer without aspirin, while 1.6 percent develop it with aspirin, for an absolute risk reduction of 0.4 percent over five years.

Another way to present these results would be to say that a woman's chance of being free from breast cancer over the next five years was 98.4 percent if she used aspirin and 98 percent if she did not. Seeing the actual risks leaves a very different impression than a statement like "aspirin lowers breast cancer risk by 20 percent." (See "Research Basics: How Big Is the Difference?")

Against What Size Harms?

Is the potential benefit of aspirin big enough to outweigh its known harms?

Unfortunately, aspirin, like most drugs, can have side effects. These, according to the U.S. Preventive Services Task Force, include a small risk of serious (and possibly fatal) bleeding in the stomach or intestine, or strokes from bleeding in the brain -- harms briefly noted but not quantified in the original study or in most media reports. To decide whether aspirin is worth taking, women need to know how the potential size of aspirin's benefit in reducing breast cancer compares with the drug's potential harms.

Sound medical practice dictates doing the same kind of calculation -- of potential benefits against potential harms -- anytime you consider taking a drug.

We provide the relevant information in the "Aspirin Study Facts," below. The first column shows the health outcome being considered (e.g., getting breast cancer, having a major bleeding event). The second column shows the chance of the outcome over five years for women *not* taking aspirin. The third column shows the corresponding chance for women taking aspirin. And the fourth column shows the difference -- the possible effect of aspirin.

As the table shows, the size of the known risk for stomach bleeding to a woman taking aspirin daily nearly matches the size of the still-hypothetical benefit in terms of breast cancer protection. That kind of comparison might lead some women to conclude that the tradeoff doesn't warrant the risk.

While it may take you some time to become familiar with this table, we think this sort of presentation would be helpful in many situations; for example, whenever people are deciding about taking a new medication or undergoing elective surgery.

Is It Really Aspirin?

Does aspirin really prevent breast cancer, or is there some other difference between women in the study that could account for the difference in cancer rates?

Can we be sure that aspirin was responsible for the "20 percent fewer" breast cancers that the Columbia researchers found among aspirin users compared with nonusers?

To understand why not, it is necessary to know some of the details about how the study was conducted.

The researchers collected information from all of the women in New York's Nassau and Suffolk counties on Long Island, who were diagnosed with breast cancer in 1996 and 1997. For comparison, they matched these women with others who did not have breast cancer, but who were about the same age and from the same counties. The researchers asked all the women about their use of aspirin.

They found that aspirin use was more common among the women without breast cancer. While the researchers were careful to report that the use of aspirin was "associated" with reduced risk of breast cancer, the media used stronger language, suggesting aspirin played a role in preventing breast tumors.

Unfortunately, this kind of study -- an observational study -- cannot prove that it was the aspirin that lowered breast cancer risk. Strictly speaking, the researchers demonstrated only that there is an association between aspirin and breast cancer.

Consider how an association between aspirin and breast cancer could exist even if aspirin has no effect on breast cancer.

It could be that women who use aspirin regularly are already at a lower risk of breast cancer. Imagine, for example, there was a gene that protected against breast cancer but also made people more susceptible to pain. Women who carried this gene would be more apt to use aspirin for pain relief. The lower breast cancer risk in aspirin users might simply reflect the fact that they had this gene. In other words, aspirin might have nothing to do with the findings. To really know if aspirin lowers breast cancer risk would require a different kind of study -- a randomized trial. (See "Research Basics: Cause or Association?")

Nonetheless, observational studies are important (and often crucial) in building the case for doing a randomized trial. In this instance, the researchers had a theory for how aspirin might prevent breast cancers. They predicted that it would only be true for certain kinds of cancers (so-called hormone receptor positive cancers, the most dangerous kind, which account for about 60 percent of all breast cancers). And that is just what they observed: The association between aspirin and breast cancer was not seen in hormone receptor negative cancers. That the researchers' prediction was correct supports (but does not prove) the idea that aspirin reduces risk. The next logical step would be a randomized trial.

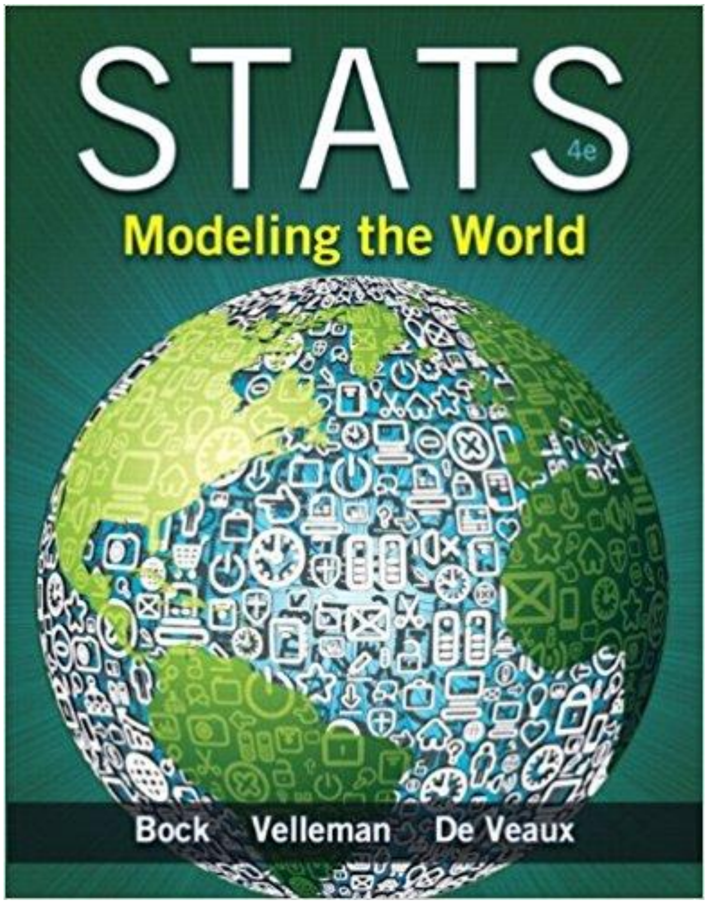
The difference between "cause" and "association" may seem subtle, but it is actually profound. Even so, people -- like the headline writers in this case -- often go beyond the evidence at hand and assume that an association is causal. Readers should know that many associations do not reflect cause and effect.

The Bottom Line

In a large observational study, researchers found slightly fewer breast cancers among women who took aspirin regularly compared with women who did not. Because aspirin's benefit in reducing breast cancer (assuming it can be proven) was small, it may not outweigh the drug's known harms. While it is possible that aspirin itself reduces the risk of breast cancer, we cannot be sure from this study. It would take a randomized trial to be certain. Fortunately, one has just been completed by researchers at Harvard Medical School, and the results are expected in the very near future. Until then, it is too soon to recommend taking aspirin to prevent breast cancer.

AP Statistics Shopping list:

1. *TEXTBOOK: STATS Modeling the World: 4th Edition*



Note: Try looking on Amazon or other websites for this book. It will be pricey if you buy it off of the pearson website.

2. Graphing Calculators: Required for in-class and homework (TI-84 plus).

You will not be able to pass this course without a graphing calculator. The course is created around the TI-84 Plus.