ehp

LESSON:

What's Your Hypothesis?

Summary: Students read the article "In the Thick of Air Pollution" and review air pollution data and cardiovascular disease death rates to develop a hypothesis about their relationship. Then, students test their hypothesis using statistical analysis (i.e., Pearson's r) to examine whether a correlation exists between air pollution and cardiovascular disease. Finally, students develop a new hypothesis that scientists could research to learn more about how air pollution affects the cardiovascular system. This lesson extends the discussion of a topic addressed within an article in the *EHP Student Edition*.

EHP Article: "In the Thick of Air Pollution," *EHP Student Edition*, May 2005: A116. http://ehp.niehs.nih.gov/docs/2005/113-2/ss.html

Objectives: By the end of this lesson students should be able to:

1. describe how air pollution, including fine particles, affects human health;
2. identify at least three components of air pollution;
3. calculate descriptive statistics, such as mean, median, and range;
4. calculate correlational statistics, such as Pearson's r;
5. draw conclusions about whether a correlation between different components of air pollution and cardiovascular disease exists; and
6. write a hypothesis to study how air pollution affects the cardiovascular system.

Class Time: 2–3 hours (allow more time if students are doing calculations manually or assign the calculations as homework)

Grade Level: 11–12

Subjects Addressed: Biology, Life Science, Environmental Science, Math, Statistics, Health

Prepping the Lesson (15–20 minutes)

INSTRUCTIONS:

1. Obtain a class set of *EHP Student Edition*, May 2005, or download article at http://ehp.niehs.nih.gov/docs/2005/113-2/ss.html and make copies.
2. Review the article "In the Thick of Air Pollution."
3. Review the student instructions and decide whether you want students to perform the Pearson's r calculations manually (with calculators), or by computer using Microsoft Excel spreadsheet software.
4. Make copies of the student instructions.
5. If students will perform the calculations by computer, download the Microsoft Excel file "calcpearsonsrtemplate_hypo.xls" at http://www.ehponline.org/science-ed and review how to use the template.

MATERIALS (per student):

- 1 copy of *EHP Student Edition*, May 2005, or 1 copy of the article "In the Thick of Air Pollution"
- 1 copy of the student instructions
- At least 3 pieces of graph paper
- Scientific calculator (calculators should be able to square numbers and have a minimum of 10 digits), if students will perform the calculations manually
- Computer with spreadsheet software Microsoft Excel and the file "calcpearsonsrtemplate_hypo.xls", if students will perform the calculations by computer

VOCABULARY:

Air pollution

Atherosclerosis
Benzene
Cardiovascular disease
Coefficient of correlation
Correlational statistics
Criteria air pollutants
Descriptive statistics
Fine particles
Hazardous air pollutants
Hypothesis
Micrograms per cubic meter (mg/m^3 or $\mu g/m^3$)
Parts per billion carbon (ppbC)
Pearson's r
$PM_{2.5}$
PM_{10}
Sample size (N)
Sum (Σ)
Toxic air pollutants

BACKGROUND INFORMATION:

Background information is included in the student instructions. Additional resources about air pollution and cardiovascular disease are provided below. The statistical analysis conducted in this lesson has a number of limitations. First, the air pollution data for each contaminant used in the lesson represent the mean value of all the monitoring stations located in that state. Since monitoring stations are not randomly placed in the state, the calculated mean of the monitoring station means is not a true average of the entire state. Second, cardiovascular disease develops over many years. Comparing the exposure concentration of a contaminant in one year with the mortality rate for cardiovascular disease in one year does not consider the longer time period involved in the development of the disease.

RESOURCES:

Environmental Health Perspectives, Environews by Topic page. Choose Air Pollution, Environmental Disease, Particulate Matter, Science Selections, http://ehp.niehs.nih.gov/topic

American Heart Association, Search "air pollution" on main website, http://www.heart.org/

American Health Association "Air Pollution and Cardiovascular Disease: A Statement for Healthcare Professionals from the Expert Panel on Population and Prevention Science of the American Heart Association," Circulation, 2004; 109(21):2655–2671, http://circ.ahajournals.org/cgi/content/full/109/21/2655

National Institute of Environmental Health Sciences, Highlights in Environmental Health Sciences Research, "Long-term Exposure to Particulate Matter and Risk of Death from Heart Disease," http://www.niehs.nih.gov/dert/profiles/hilites/2003/pm-heart.htm

National Institute of Environmental Health Sciences, "Exposure to Particulate Matter and Other Air Pollutants Decrease Cardiac Autonomic Control", http://www.niehs.nih.gov/dert/profiles/hilites/2004/cardiac.htm

National Institute of Environmental Health Sciences, The Role of Environmental Agents in Cardiovascular Disease, 2002 Workshop Summary, http://www.niehs.nih.gov/dert/events/cardio02/report.pdf

National Library of Medicine, Specialized Information Services (SIS) Division, Internet Resources for Outdoor Air Pollution, http://www.sis.nlm.nih.gov/Tox/outdoorair.html

U.S. Environmental Protection Agency, Search "cardiovascular disease" on main website, http://www.epa.gov/

Implementing the Lesson

INSTRUCTIONS:

1. Hand out copies of the *EHP Student Edition*, May 2005, and refer your students to the article "In the Thick of Air Pollution," or hand out copies of the article.
2. Hand out the student instructions.
3. Lead a discussion about the research described in the article, focusing on what was studied, what the hypothesis was, and what the conclusions were (student instructions step 1). Be sure to explain how atherosclerosis impacts cardiovascular disease, a group of diseases of the heart and blood vessels. Atherosclerosis is the buildup of a fatty/wax-like substance (plaque) on the inside of the arterial walls narrowing the passageway for blood to travel.

Atherosclerosis is the main cause of cardiovascular disease.

4. Tell students that they will be reviewing additional data about air pollution and cardiovascular disease to see if these data support the conclusions of the article. Have the students complete the student instructions, steps 2–7 to examine whether a correlation exists between air pollution and cardiovascular disease. Be sure to have students round the data from the "Background Reading & Data: Air Pollution and Cardiovascular Disease" handout to the nearest hundredth decimal or to your preference. Also, give students notice as to which states have air pollution above the standard, as noted on the "Background Reading & Data: Air Pollution and Cardiovascular Disease" handout. The EPA air quality standard for $PM_{2.5}$ is 15 $\mu g/m^3$ (annual mean). The standard for PM_{10} is 50 $\mu g/m^3$ (annual mean). There is no EPA air quality standard for benzene as there is no identified safe level of exposure.
5. If you are having students perform the calculations using a calculator, review the student handout "Calculating Correlational Statistics (Pearson's r)." If you are having students perform the calculations using a computer, provide access to computers with spreadsheet software, such as Microsoft Excel. It is recommended that the spreadsheet file "calculating Pearsonsr template.xls" be downloaded onto each computer ahead of time.
6. Review, as needed, the formula for calculating Pearson's r and the terms represented in the formula (e.g., Σ means to sum or add; N means the total number of samples, such as 10 states or 50 states).
7. Have students present their final hypotheses and conclusions to the class (student instructions, step 7). Tell students that correlation only shows a relationship, it does not show cause and effect.

NOTES & HELPFUL HINTS:

Instructions for how the data provided in the "Background Reading & Data: Air Pollution and Cardiovascular Disease" handout were obtained are outlined below.

Air Pollution Data:

- Go to the Air Data website at the Office of Air & Radiation, U.S. Environmental Protection Agency: http://www.epa.gov/air/data/index.html
- Click "Reports and Maps"
- Click "Select Geographic Area"
- Click the states you want included in the report, then click "Go"
- For $PM_{2.5}$ and PM_{10} data, click "Monitor Values" under Criteria Air Pollutants, or for benzene data, click "Monitor Values" under Hazardous Air Pollutants
- Choose the pollutant(s) and year(s) you want included in the report, then click "Generate Report"
- A very detailed report will be generated. (For purposes of this lesson, an average of the annual means for each monitoring station was calculated.)

Cardiovascular Disease Death Rates:

- Go to the American Heart Association website: http://www.heart.org/
- Click "Publications and Resources"
- Click "Statistics"
- Under "Heart Disease and Stroke Statistics – 2005 Update," click "More"
- Click "Heart Disease and Stroke Statistics – 2005 Update" again
- Click "Heart Disease and Stroke Statistics – 2005 Update" again for the pdf file of the report
- After the pdf has opened, the cardiovascular disease rates can be found on page 9 of the document

- You could assign this activity as homework.
- This activity could complement a lesson on the cardiovascular system.
- You could have students work in groups, each group determining the Pearson's r and r^2 for one type of air pollution and cardiovascular disease, then have groups present their findings.
- Depending upon which 10 states are selected for the analysis, your students' answers will vary. Answers may even contradict what is expected, as not all variations with 10 states were tested. If results are different than what is expected, refer to results for the entire set of data, including all 50 states, the District of Columbia, and Puerto Rico (see "Assessing the Lesson," below). In addition, if a state is chosen that does not have benzene data available, you will need to adjust the sample size (N) in your calculations so that this state is not included.
- If you have your students use computers with Microsoft Excel, the function for calculating Pearson's r is: PEARSON(Array 1, Array 2) where Array 1 is the first set of data you want to compare and Array 2 is the second set of data. For example, if you use the provided Microsoft Excel file (calculating pearsonsr template.xls), the function is PEARSON(C4:C13,F4:F13) for comparing $PM_{2.5}$ and cardiovascular disease.
- You could also use Microsoft Excel to graph the data highlighting all three parameters of interest (i.e., state, PM, and

cardiovascular disease), click "Insert" on the Tool Bar, select "In Chart," click the "Custom Types" tab, and highlight "Lines on 2 Axes."
- As a challenge, have students calculate Pearson's r for the entire set of data, including all 50 states, the District of Columbia, and Puerto Rico.
- As extension activities, have students design an experiment to test their hypothesis, or calculate r^2, also known as the coefficient of determination. To calculate r^2, square the coefficient of correlation, or Pearson's r. (The coefficient of determination, ranging from 0.00 to 1.00, can also be helpful in understanding the relationship between two variables. This number explains how much of the variability of one variable can be accounted for by being related to the second variable. Convert the coefficient of determination to a percentage by multiplying by 100. This percentage will indicate the amount of variation in cardiovascular disease that is due to the variation in that component of air pollution.)

Aligning with Standards

SKILLS USED OR DEVELOPED:
- Communication (note taking, oral, written—including summarization)
- Comprehension (reading)
- Computation
- Critical thinking and response
- Graphing

SPECIFIC CONTENT ADDRESSED:
Descriptive statistics, correlational statistics, cardiovascular disease, air pollution

NATIONAL SCIENCE EDUCATION CONTENT STANDARDS MET:

Unifying Concepts and Processes
- Evidence, models, and explanation

Science as Inquiry
- Abilities necessary to do scientific inquiry
- Understanding about scientific inquiry

Science in Personal and Social Perspectives
- Personal and community health
- Environmental quality
- Natural and human-induced hazards

Assessing the Lesson

Students provide written responses to the questions on the student instructions. Examples of complete student responses are outlined below for each question. Please note: your students' answers will vary.

Step 1:

a) What was studied?
Air pollution and cardiovascular disease or atherosclerosis, or the relationship between exposure to fine particles and the thickening of the carotid artery

b) What was the hypothesis?
Exposure to fine particles affects cardiovascular disease, or increased exposure to fine particles will thicken the arteries or cause atherosclerosis

c) What were the conclusions?
1) Increase in exposure to fine particles resulted in more thickening of the carotid artery, 2) exposure to ozone was not linked to atherosclerosis, 3) more research is needed with at-risk populations and more subjects, taking indoor air pollution into account

Step 2:

If the conclusions in the article are correct, what is your hypothesis about the relationship between the three components of air pollution ($PM_{2.5}$, PM_{10}, and benzene) and cardiovascular disease? Explain.

Students should hypothesize that $PM_{2.5}$ would affect cardiovascular disease based on the article. There is no information to indicate PM_{10} or benzene would affect cardiovascular disease.

Step 3:

Depending upon the states chosen by your students, the values for the mean, median, minimum value, maximum value, and range will vary. The table below provides the descriptive statistics for the entire set of data, including all 50 states, the District of Columbia, and Puerto Rico.

Descriptive Statistics for All Data, Including the 50 States, the District of Columbia, and Puerto Rico

		Air Pollution			Cardiovascular Disease
		$PM_{2.5}$ ($\mu g/m^3$)	**PM_{10}** ($\mu g/m^3$)	**Benzene** (ppbC)	**Death Rate** (per 100,000 population)
Descriptive Statistics	Total Sum	607.3062	1231.5802	132.1345	16621.2
	Mean	11.6790	23.6842	2.7528	328.1
	Median	11.6556	23.3333	2.5906	313.9
	Minimum Value	4.6200	15.6667	0.1650	247.5
	Maximum Value	16.5565	39.4546	8.6500	424.2
	Range	11.9365	23.7879	8.4850	176.7

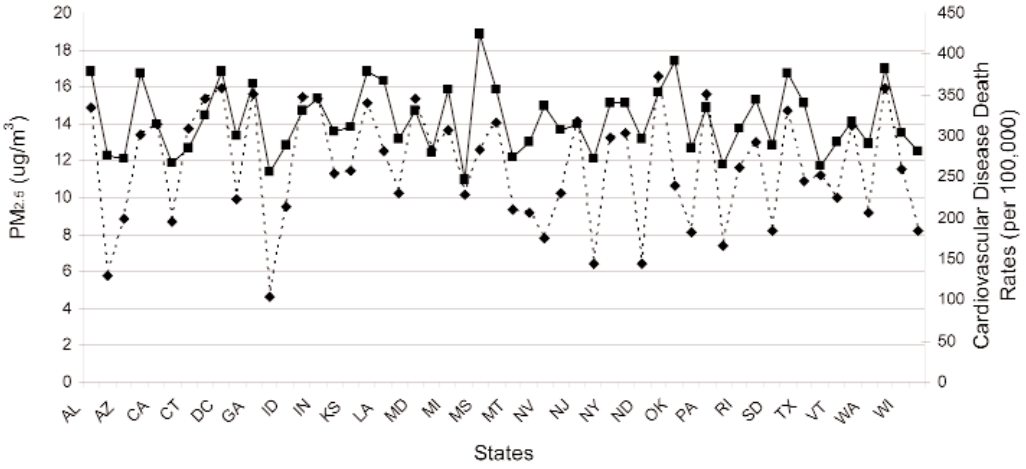
Step 4:

Depending upon the states chosen by your students, the graphs will vary. Below are graphs plotting the entire set of data including all 50 states, the District of Columbia, and Puerto Rico. (Cardiovascular disease is the solid line in each graph.

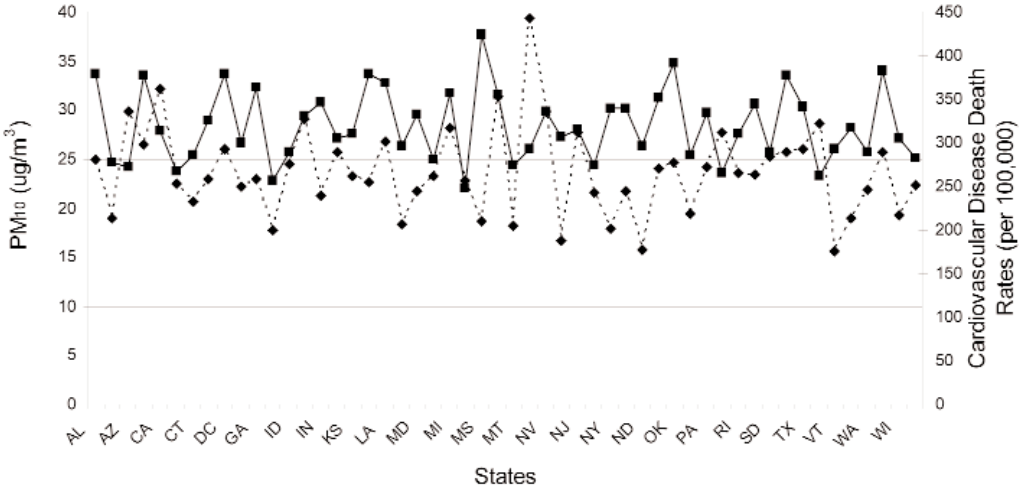
What are your conclusions by looking at the graphs?

Depending upon the states chosen, students may note the line for $PM_{2.5}$ appears to be the most similar to, or correlated with, the cardiovascular disease line. PM_{10} is somewhat similar to the line for cardiovascular disease and benzene is the least similar, or the least correlated of the three.

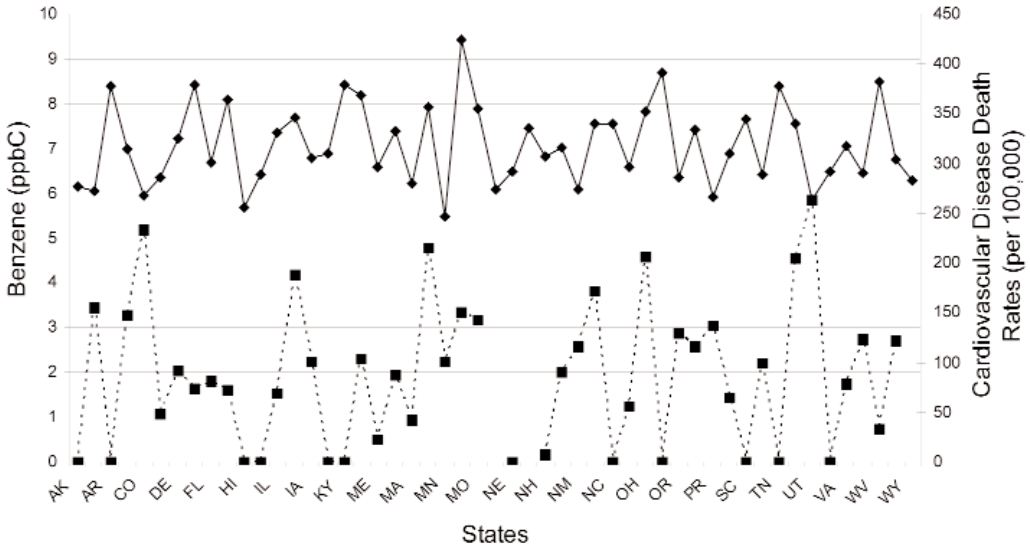
$PM_{2.5}$ and Cardiovascular Disease



PM_{10} and Cardiovascular Disease



Benzene and Cardiovascular Disease



Step 5: Students calculate the coefficient of correlation using a calculator or the Excel spread sheet.
Step 6: Depending upon the states chosen by your students, the coefficients of correlation (Pearson's r) will vary. Provided that:

	Coefficient of Correlation (r)	Type of Correlation	Size of Correlation
$PM_{2.5}$ and cardiovascular disease:	0.6828	Positive	Large
PM_{10} and cardiovascular disease:	0.1450	Positive	Small
Benzene and cardiovascular disease:	0.0047	Positive	Small

Below are the Pearson's r for all of the data:
Size of Correlations:
- 0.00 No correlation
- 0.01–0.29 Small
- 0.30–0.49 Medium
- 0.49–1.0 Large

Depending upon which 10 states are selected for the analysis, answers may even contradict what is expected, as not all variations of 10 states were tested. If results are different than what is expected, refer to results for the entire set of data, including all 50 states, the District of Columbia, and Puerto Rico.

Step 7:

a) How do the results of your statistical analysis compare to your hypothesis in step 2 and the hypothesis from the article?
Most likely, students will find that their hypothesis linking $PM_{2.5}$ and cardiovascular disease will hold true and was similar to the article's conclusion. Depending upon the states chosen, students may conclude differently. When the entire set of data is used in the correlation, $PM_{2.5}$ is found to be correlated with cardiovascular disease while PM_{10} and benzene are not.
b) Write a new hypothesis that scientists could research to learn more about how air pollution affects the cardiovascular system.
Students could suggest hypotheses linking air pollution and cardiovascular disease in at-risk populations, linking indoor air pollution to cardiovascular disease, or linking air pollution to specific cardiovascular diseases, such as heart attacks or stroke.

Authors and Reviewers

AUTHORS: Laura Hemminger and Barry Schlegel, University of Medicine and Dentistry of New Jersey–School of Public Health

REVIEWERS: Susan Booker, Stefani Hines, Liam O'Fallon, Kimberly Thigpen Tart

ehp

Student Instructions: What's Your Hypothesis?

Purpose: To develop and test your own hypothesis about whether air pollution and cardiovascular disease are related.

Step 1: Read the article "In the Thick of Air Pollution," *EHP Student Edition,* May 2005, p. A116, and answer the questions below.

a) What was studied?

b) What was the hypothesis?

c) What were the conclusions?

d) What are the limitations?

Step 2: Compare additional data about air pollution and cardiovascular disease by reviewing the "Background Reading and Data: Air Pollution and Cardiovascular Disease" handout to see if these new data support the conclusions in the article.

a) If the conclusions in the article are correct, what is your hypothesis about the relationship between three components of air pollution ($PM_{2.5}$, PM_{10}, and benzene) and cardiovascular disease? *(Remember: a hypothesis is a statement that proposes a possible explanation to some phenomenon or event. A useful hypothesis is a testable statement which includes a tentative relationship rather than just a simple prediction.)* Explain.

Step 3: Test your hypothesis by conducting a statistical analysis. First, locate the data for your state and nine surrounding states on the "Background Reading & Data: Air Pollution and Cardiovascular Disease" handout. Then either:

a) enter the data into a Microsoft Excel spreadsheet provided by your teacher; OR
b) enter the data into the table below and calculate the descriptive statistics. Round the air pollution data to the nearest hundredth decimal or as instructed by your teacher before entering the data into the table. *(Descriptive statistics describes a set of data, such as its central tendencies and how it is distributed—mean, median, etc.)*

Be sure to note which states have air pollution data above the standard. The EPA air quality standard for $PM_{2.5}$ is 15 $\mu g/m^3$ (annual mean). The standard for PM_{10} is 50 $\mu g/m^3$ (annual mean). There is no EPA air quality standard for benzene as there is no identified safe level of exposure.

State		Air Pollution			Cardiovascular Disease Death Rate (per 100,000 population)
		$PM_{2.5}$ ($\mu g/m^3$)	PM_{10} ($\mu g/m^3$)	Benzene (ppbC)	
1)					
2)					
3)					
4)					
5)					
6)					
7)					
8)					
9)					
10)					
Descriptive Statistics	Total Sum				
	Mean				
	Median				
	Minimum Value				
	Maximum Value				
	Range				

Step 4: On graph paper, plot the data for your 10 states using the scatterplot method. Create three graphs comparing each component of air pollution to cardiovascular disease. Plot the states on the 'X' axis, one component of air pollution data on the 'Y' axis, and cardiovascular disease death rates on the secondary 'Y' axis. (The two 'Y' axes will have different scales.) What are your conclusions by looking at the graphs?

Step 5: Now use correlational statistics to determine whether there is a relationship (or correlation) between each type of air pollution and cardiovascular disease. *(Correlational statistics is a method to better show if two sets of data are related, or correlated. However, correlation does not mean causation—that one variable causes the other.)*

Calculate the **coefficient of correlation**. Use either the "Calculating Correlational Statistics (Pearson's r)" handout to manually calculate the coefficients of correlation, or spreadsheet software such as Microsoft Excel.

(The coefficient of correlation will range from –1.00 to +1.00. If the number is positive, the correlation is said to be a positive correlation or a direct relationship. That is, when scores for one variable increase/decrease, scores for a second variable do the same. In a negative ,or inverse, correlation, noted by a negative number, scores for two variables act opposite of one another; as one increases the other decreases or vice versa. If the correlation was a perfect +1.00 or -1.00, then we could accurately predict a score for one variable from a score for the other variable. However, correlations are rarely perfect. We can usually predict scores only with a certain degree of probability. The closer the correlation is to ∀1.00, the greater our probability, or chance, of being correct. The strength of a correlation depends on its size, not its sign. For example, a correlation of –.72 is stronger than a correlation of +.53. Another way to think of this is if the correlation is +.1, then there is only a 10% chance that the two variables are related. If the correlation is +.72, then there is a 72% chance the variables are related or correlated.)

The most commonly used coefficient of correlation is the Pearson's product-moment correlation (Pearson's r), named for the English statistician Karl Pearson. It is a measure of the degree to which pairs of scores occupy the same relative position in each of their data sets. One of the formulae for calculating Pearson's r is:

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right)}}$$

where X and Y represent the variables you are examining to determine if there is a correlation.

Step 6: What are the coefficients of correlation (Pearson's r) when you compare each type of air pollution with cardiovascular disease? Include the type of the correlation (positive or negative) and the size of the correlation (small, medium, or large—see range below).

	Coefficient of Correlation (r)	Type of Correlation	Size of Correlation
$PM_{2.5}$ and cardiovascular disease:			
PM_{10} and cardiovascular disease:			
Benzene and cardiovascular disease:			

Size of Correlations:

- 0.00 No correlation
- 0.01–0.29 Small
- 0.30–0.49 Medium
- 0.49–1.0 Large

Step 7: Draw your own conclusions about whether air pollution affects the cardiovascular system and answer the questions below.

a) How do the results of your statistical analysis compare to your hypothesis in step 2 and the hypothesis from the article?

b) Write a new hypothesis that scientists could research to learn more about how air pollution affects the cardiovascular system.

Background Reading and Data: Air Pollution and Cardiovascular Disease

Air pollution affects everyone, everywhere, from cities to farms and from the mountains to the shores. In part, air pollution is created from burning coal, oil, and other fossil fuels, as well as manufacturing chemicals. However, air pollution is not only caused by factories. Each person contributes to air pollution through everyday activities, such as driving your car or truck, having your clothes dry cleaned, and smoking. As the air pollutants produced by daily living and working accumulate in the air, they eventually achieve high enough concentrations to harm public health and the environment. Some people may experience health problems–including burning eyes, an irritated throat, or breathing difficulties—when concentrations are high enough. Air pollution can damage the immune, neurological, respiratory, and reproductive systems, and cause cancer, and even death with long-term exposure.

The U.S. Environmental Protection Agency (EPA), through the Clean Air Act, monitors *criteria* and *toxic* air pollutants across the country. The EPA has established air quality standards for the six criteria air pollutants: carbon monoxide, lead, nitrogen dioxide, ozone, particulate matter, and sulfur dioxide. These are common pollutants, and their monitoring is mandated by law.

> Particulate matter, or PM, is a *criteria air pollutant* that comprises particles found in the air, including dust, dirt, soot, smoke, and liquid droplets. Particles can be large or dark enough to be seen, or they can be so tiny that they can only be seen through an electron microscope. PM can be emitted directly from various sources, such as cars, trucks, construction sites, and burning of wood, as well as formed in the atmosphere when emissions of nitrogen oxides, sulfur oxides, ammonia, organic compounds, and other gases react in the air. Research studies have associated exposure to elevated levels of these particles in the air with damaging health effects. PM is divided into two categories: $PM_{2.5}$ and PM_{10}. $PM_{2.5}$, or "fine particles," are smaller than 2.5 μm (micrometers) in diameter (approximately 1/30th the average width of a human hair). $PM_{2.5}$ are very tiny particles that may pose the greatest health risk because they can reach deep into the lungs. (These fine particles were studied in the article "In the Thick of Air Pollution.") The EPA air quality standard for $PM_{2.5}$ is 15 $\mu g/m^3$ (annual mean). PM_{10}, or "coarse particles," are less than 10 μm in diameter and can be inhaled into and accumulate in the respiratory system. The standard for PM_{10} is 50 $\mu g/m^3$ (annual mean).

The EPA also monitors the release of 188 toxic air pollutants, also known as hazardous air pollutants. These pollutants are known or suspected to cause cancer or other serious health effects, such as reproductive effects or birth defects, or adverse environmental effects. Monitoring of toxic air pollutants is not mandated by the Clean Air Act.

> Benzene is one example of a *toxic air pollutant*. Benzene is a clear, colorless, highly flammable liquid that has a distinctive smell. Benzene is widely used in the manufacture of detergents and pharmaceuticals, and is found in gasoline, and even tobacco smoke. Benzene released from cigarettes accounts for almost half the national exposure to benzene. Short-term exposure to benzene can cause nervous system disorders, immune system depression, and anemia while long-term exposure can cause cancer. There is no EPA standard for benzene because there is no identified safe level of exposure.

The table below contains the 2001 air pollution data for $PM_{2.5}$, $PM_{10,}$ and benzene, and the cardiovascular disease death rates for the same year. The air pollution data represents an average for that state or location. Data for 2001 were used as this was the latest year where both air pollution data and cardiovascular disease death rates were available.

2001 Air Pollution Data and Cardiovascular Disease Death Rates for the United States

State	Air Pollutant			Cardiovascular Death Rates (per 100,000 population)
	$PM_{2.5}$ ($\mu g/m^3$)	PM_{10} ($\mu g/m^3$)	Benzene* (ppbC)	
U.S. Standard	15.0000	50.0000	———	———
AL	14.8880	24.9470	5.6389§	379.4
AK	5.7667	19.0000	N/A•	276.8
AZ	8.8786	29.8235	3.4300	272.6
AR	13.4097	26.5000	2.7450§	377.5
CA	14.0000	32.1191	3.2789	313.9
CO	8.6857	22.5714	5.1925	267.9
CT	13.7706	20.7368	1.0610	285.5
DE	15.4000	23.0000	2.0400	325.6
DC	15.9400	26.0000	1.6330	379.0
FL	9.8844	22.2131	1.8217	300.8
GA	15.5852	23.0556	1.6025	364.0
HI	4.6200	17.8333	4.2680§	256.2
ID	9.4905	24.4846	3.0400§	289.4
IL	15.4114	29.0909	1.5250	331.3
IN	15.3300	21.2826	4.1874	346.1
IA	11.3353	25.7826	2.2258	305.7
KS	11.4706	23.2308	8.6500§	310.1
KY	15.1609	22.6250	4.0660§	379.0
LA	12.4923	26.8333	2.3060	368.5
ME	10.2737	18.4118	0.4968	296.8
MD	15.3273	21.7273	1.9290	332.0
MA	12.5962	23.3333	0.9240	280.2
MI	13.6735	28.1539	4.7942	357.0
MN	10.2000	22.8889	2.2410	247.5
MS	12.6235	18.6667	3.3350	424.2
MO	14.0957	31.3529	3.1675	355.9

*2001 benzene air pollution data not available for all states.
§ Data used were from the year closest to 2001 (1994 to 2003), depending upon the state and available data.
• Data not available for any year.
Sources: Air pollution data: U.S. Environmental Protection Agency. 2001. Monitor Value Report—Criteria Air Pollutants and Hazardous Air Pollutants. (Website: http://www.epa.gov/air/data/)
Cardiovascular disease death rates: American Heart Association. 2004. Heart Disease and Stroke Statistics—2005 Update. Dallas, TX: American Heart Association. (Website: http://www.heart.org/)

2001 Air Pollution Data and Cardiovascular Disease Death Rates for the United States

State	Air Pollutant			Cardiovascular Death Rates (per 100,000 population)
	$PM_{2.5}$ ($\mu g/m^3$)	PM_{10} ($\mu g/m^3$)	Benzene* (ppbC)	
U.S. Standard	15.0000	50.0000	———	———
MT	9.3789	18.2439	N/A•	274.4
NE	9.1589	39.4546	2.1500§	292.7
NV	7.8300	29.8621	N/A•	336.3
NH	10.2375	16.6471	0.1650	307.2
NJ	14.1292	27.7778	2.0190	315.5
NM	6.4278	21.5667	2.5725	273.4
NY	13.2122	17.9333	3.8243	340.0
NC	13.4702	21.8000	0.3000§	339.5
ND	6.4250	15.7500	1.2450	296.0
OH	16.5565	24.0260	4.5957	352.5
OK	10.6600	24.6154	2.9200§	391.6
OR	8.0964	19.5000	2.8650	285.8
PA	15.6257	24.2653	2.5911	334.0
PR	7.3800	27.7000	3.0475	265.8
RI	11.6556	23.5833	1.4253	309.5
SC	13.0435	23.3889	0.6940§	344.1
SD	8.1727	25.3077	2.2100	288.9
TN	14.7233	25.7586	2.5900§	377.7
TX	10.8960	26.0930	4.5422	340.8
UT	11.2294	28.7000	5.8600	262.7
VT	9.9667	15.6667	2.9513§	292.6
VA	13.9150	18.9444	1.7554	317.5
WA	9.1514	21.8846	2.7457	290.3
WV	15.9000	25.7857	0.7500	382.6
WI	11.5445	19.2857	2.7163	304.4
WY	8.2100	22.3750	N/A•	282.5

*2001 benzene air pollution data not available for all states.
§ Data used were from the year closest to 2001 (1994 to 2003), depending upon the state and available data.
• Data not available for any year.
Sources: Air pollution data: U.S. Environmental Protection Agency. 2001. Monitor Value Report—Criteria Air Pollutants and Hazardous Air Pollutants. (Website: http://www.epa.gov/air/data/)
Cardiovascular disease death rates: American Heart Association. 2004. Heart Disease and Stroke Statistics—2005 Update. Dallas, TX: American Heart Association. (Website: http://www.heart.org/)

Calculating Correlational Statistics (Pearson's r)

- Use Tables 1, 2, and 3, provided below, to calculate the coefficient of correlation, or Pearson's r, between each component of air pollution and cardiovascular disease.
- Values for X can be found on the student handout table titled "2001 Air Pollution Data and Cardiovascular Disease Death Rates for the United States."
- Values for Y, Y2, and XY are provided in Table 4, at the end of this handout to assist in your calculations. (Be sure to round to the nearest hundredth decimal, or as instructed by your teacher.)

Table 1: The Correlation of $PM_{2.5}$ and Cardiovascular Disease

State	$PM_{2.5}$ ($\mu g/m^3$)		Cardiovascular Disease Death Rates (per 100,000 population)		
	X	**X^2**	**Y**	**Y^2**	**XY**
Total	**ΣX**	**ΣX^2**	**ΣY**	**ΣY^2**	**ΣXY**
Sample Number (N)		N = number of samples (e.g., 10 if you are using 10 states)			

Use the space below to show your work.

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right)}}$$

National Institute of Environmental Health Sciences
Community Outreach & Education Program

Table 2: The Correlation of PM_{10} and Cardiovascular Disease

State	PM_{10} ($\mu g/m^3$)		Cardiovascular Disease Death Rates (per 100,000 population)		
	X	**X^2**	**Y**	**Y^2**	**XY**
Total	**ΣX**	**ΣX^2**	**ΣY**	**ΣY^2**	**ΣXY**
Sample Number (N)		N = number of samples (e.g., 10 if you are using 10 states)			

Use the space below to show your work:

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right)}}$$

Table 3: The Correlation of Benzene and Cardiovascular Disease

State	Benzene (ppbC)		Cardiovascular Disease Death Rates (per 100,000 population)		
	X	**X^2**	**Y**	**Y^2**	**XY**
Total	**ΣX**	**ΣX^2**	**ΣY**	**ΣY^2**	**ΣXY**
Sample Number (N)		N = number of samples (e.g., 10 if you are using 10 states) NOTE: Your sample size will be limited by the number of benzene samples you have. If two of the states you selected do not have benzene data, then your N will equal 8, not 10.			

Use the space below to show your work:

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right)}}$$

Table 4: Values for Y, Y^2, and XY

	Cardiovascular Disease Death Rate (per 100,000 population)		Cardiovascular Disease Death Rate Multiplied By:		
			$PM_{2.5}$ ($\mu g/m^3$)	PM_{10} ($\mu g/m^3$)	Benzene (ppbC)
	Y	Y^2	XY	XY	XY
AL	379.40	143944.36	5648.5072	9464.8918	2139.3987
AK	276.80	76618.24	1596.2226	5259.2000	N/A
AZ	272.60	74310.76	2420.3064	8129.8861	935.0180
AR	377.50	142506.25	5062.1618	10003.7500	1036.2375
CA	313.90	98533.21	4394.6000	10082.1855	1029.2467
CO	267.90	71770.41	2326.8990	6046.8781	1391.0708
CT	285.50	81510.25	3931.5063	5920.3564	302.9155
DE	325.60	106015.36	5014.2400	7488.8000	664.2240
DC	379.00	143641.00	6041.2600	9854.0000	618.9070
FL	300.80	90480.64	2973.2275	6681.7005	547.96734
GA	364.00	132496.00	5673.0128	8392.2384	583.3100
HI	256.20	65638.44	1183.6440	4568.8915	1093.4616
ID	289.40	83752.36	2746.5507	7085.8432	879.7760
IL	331.30	109759.69	5105.7968	9637.8152	505.2325
IN	346.10	119785.21	5305.7130	7365.9079	1449.2591
IA	305.70	93452.49	3465.2012	7881.7408	680.4271
KS	310.10	96162.01	3557.0331	7203.8711	2682.3650
KY	379.00	143641.00	5745.9811	8574.8750	1541.0140
LA	368.50	135792.25	4603.4126	9888.0711	849.7610
ME	296.80	88090.24	3049.2342	5464.6222	147.4502
MD	332.00	110224.00	5088.6636	7213.4636	640.4280
MA	280.20	78512.04	3529.4552	6537.9907	258.9048
MI	357.00	127449.00	4881.4395	10050.9423	1711.5294
MN	247.50	61256.25	2524.5000	5665.0028	554.6475
MS	424.20	179945.64	5354.8887	7918.4141	1414.7070
MO	355.90	126664.81	5016.6596	11158.4971	1127.3133
MT	274.40	75295.36	2573.5702	5006.1262	N/A
NE	292.70	85673.29	2680.8100	11548.3614	629.3050
NV	336.30	113097.69	2633.2290	10042.6242	N/A

Table 4: Values for Y, Y^2, and XY

	Cardiovascular Disease Death Rate (per 100,000 population)		Cardiovascular Disease Death Rate Multiplied By:		
			$PM_{2.5}$ ($\mu g/m^3$)	PM_{10} ($\mu g/m^3$)	Benzene (ppbC)
	Y	Y^2	XY	XY	XY
NH	307.20	94371.84	3144.9600	5113.9891	50.6880
NJ	315.50	99540.25	4457.7626	8763.8959	636.9945
NM	273.40	74747.56	1757.3605	5896.3358	703.3215
NY	340.00	115600.00	4492.1480	6097.3220	1300.2620
NC	339.50	115260.25	4573.1329	7401.1000	101.8500
ND	2960	87616.00	1901.8000	4662.0000	368.5200
OH	352.50	124256.25	5836.1663	8469.1650	1619.9843
OK	391.60	153350.56	4174.4560	9639.3906	1143.4720
OR	285.80	81681.64	2313.9511	5573.1000	818.8170
PA	334.00	111556.00	5218.9838	8104.6102	865.4274
PR	265.80	70649.64	1961.6040	7362.6600	810.0255
RI	309.50	95790.25	3607.4082	7299.0314	441.1304
SC	344.10	118404.81	4488.2684	8048.1205	238.8054
SD	288.90	83463.21	2361.0930	7311.3945	638.4690
TN	377.70	142657.29	5560.9904	9729.0232	978.2430
TX	340.80	116144.64	3713.3568	8892.4944	1547.9818
UT	262.70	69011.29	2949.9634	7539.4900	1539.4220
VT	292.60	85614.76	2916.2564	4584.0764	863.5504
VA	317.50	100806.25	4418.0125	6014.8470	557.3395
WA	290.30	84274.09	2656.6514	6353.0994	797.0767
WV	382.60	146382.76	6083.3400	9865.6088	286.9500
WI	304.40	92659.36	3514.1458	5870.5671	826.8417
WY	282.50	79806.25	2319.3250	6320.9375	N/A