|  |  |
| --- | --- |
| **Title** | The Statistics of Climate Change: A real world example of graphing and data analysis |
| **Introduction** | Students will practice their skills of determining averages, graphing points, fitting lines to curves and analyzing data using a simplified version of data from the Johns Hopkins University National Morbidity and Mortality Air Pollution Study. Students will average data points of temperature and death rates, create a scatter plot of the averaged points, find a line(s) of best fit, and then use a graphing calculator to do a linear regression on the data. They will draw conclusions about what their data curves might mean to a community and how the community could plan for the future knowing that average temperatures are rising.  Background information: Scientists at Johns Hopkins University collected 14 years of daily data on temperature, climate and human health while tracking 104 cities in North America. The goal of their study was to look at how various aspects of climate change, temperature, and air pollution impact human health. In this lesson, students will use some of the actual data from the study to discover and explore the types of information that statisticians and scientists analyze. In particular, they will have the opportunity to explore how higher than normal temperatures in a city may impact death rates. As a class, students will draw conclusions about the impact of extreme temperatures on death rates using data for Chicago, Illinois. They will have the opportunity to practice their skills on another data set working in small groups. |
| **Curriculum Alignment** | **Curriculum Alignment/Essential Standard Alignment:**  Grade 8 Math Common Core Standards:  8.EE.7: Solve linear equations in one variable.  8.EE.8 c: Solve real-world and mathematical problems leading to two linear equations in two variables.  8.F.4: Construct a function to model a linear relationship between two quantities.  8.SP.1: Construct and interpret scatter plots for bivariate measurement data to investigate patters of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association and nonlinear association.  8.SP.2: Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line and informally assess the model fit by judging the closeness of the data points to the line.  Grade 8 Science Essential Standards:  8.P.2: Explain the environmental implications associated with the various methods of obtaining, managing and using energy resources. (climate change)  8.E.2: Understand the history of Earth and its life forms based on evidence of change recorded in fossil records and land forms. (climate change)  This lesson could be modified to align with High School Statistics and Earth Science standards. |
| **Learning Outcomes** | 1. Students will apply their basic knowledge of determining an average. They will use it to determine the average of 25-100 points of data. Students will explain the pros and cons of choosing various data set sizes for real world averages, showing their mastery of the concept of determining and using averages. 2. Students will apply their basic knowledge of graphing to create a scatter plot of 50 averaged points of data. Students will demonstrate their understanding and mastery of the construction and interpretation of scatter plots by creating a their own scatter plot of Atlanta temperature and death data. 3. Students will find a line(s) of best fit for the scatter plot which demonstrates their understanding of the relationship between quantitative variables. They will justify their use of one or two lines of best fit. 4. Students will use a graphing calculator to create a linear regression of the data. 5. Students will demonstrate how to interpret data and analyze their findings by making suggestions of what a community should do now to address a feature climate related health issue. |
| **Time Required and Location** | This lesson will require two to four 50 minute class sessions.  One session will be used to engage students and understand the need for data when exploring a real world problem. A second session will be used for determining averages, creating the scatter plot and finding a line(s) of best fit. Students will draw conclusions about their graphs. A third session could be used to enter the data points into the graphing calculator and to do the linear regression(s). Students doing the independent analysis of a second city would use a fourth session.  More or less time may be required depending on whether or not the students need to learn how to do input data into the graphing calculator and to do the linear regression. |
| **Materials Needed** | Handouts  Engage:  Attachment 1: News stories on heat related deaths: : Death toll from New York City heat wave rises - 2 Bronx women in their 70s's died from temperatures; Hundreds die in Indian heat wave; European heat wave caused 35,000 deaths October 10, 2003.  Attachment 2: Question sheet for students on News stories of heat related deaths.  Explore:  Computer and projector  Explain:  Graph paper, rulers, and calculators  Attachment 3a: Entire data set of temperature and death rate sorted by temperature for Chicago Illinois (could be shown with the projector)  Attachment 3b: 100 data points for temperature and death rates (class set)  Attachment 4: Averaged data points for plotting (class set)  Elaborate:  Graph paper, rulers, graphing calculator  Evaluate:  Computer access or make a class set of a copy of the article found at the listed website; Attachment 6: CHICAGO — The Windy City is preparing for a heat wave — a permanent one.  Assessment:  Attachment 5: Atlanta data  **Technology Resources:**  Each student or student pair will need a graphing calculator.  The videos and data can be shown from an internet connection using a computer and projector. |
| **Participant Prior Knowledge** | Students need to have a basic knowledge of how to determine an average of a set of numbers.  Students need to have a basic knowledge of how to plot data points to create a scatter plot.  Students also need to know what a line of best fit is.  Students also need to know the basics of using a graphing calculator to enter data points and to create a scatter plot. |
| **Facilitator Preparations** | Copy a class set of the three news articles on heat related events in Attachment 1: Death toll from New York City heat wave rises - 2 Bronx women in their 70s's died from temperatures; Hundreds die in Indian heat wave; European heat wave caused 35,000 deaths October 10, 2003. Make a class set of the student answer sheet for the news stories on heat related deaths in Attachment 2.  Copy a class set of the two data sets for averaging and for making the scatter plots shown in Attachment 3b: Chicago data sorted 100 points; Attachment 4: Averaged Chicago data.  Make a class set of the New York Times article in Attachment 6: CHICAGO — The Windy City is preparing for a heat wave — a permanent one.  Make a class set of the Atlanta data averages from Attachment 5.  Teachers need to know how to use the graphing calculator to input a data table and to create a linear regression. |
| **Activities** | **Engage:** Ask the class, what do you do when it gets really hot in the summer? You may get answers like we go swimming or we stay inside in the air conditioning or we sit around and drink lots of cold drinks or we go to the beach. Students may even answer—it’s no big deal—I just do what I do every day. Continue the discussion asking—what might happen if it stays really hot for several days or even a week? Ask the students if there are any people who might be more or less impacted by really hot weather. Write all the answers on the white board.  Let’s look at what has happened in some communities where it got really hot—hotter than normal—and stays hot for several days in a row. Have students read several short news stories about heat waves that have occurred in several cities around the world—there are three stories from newspapers in attachment 1. You could also find local stories if you live in a warm area that has had a recent heat wave. As they read, have them answer questions about the events including in attachment 2:  What happened?  When did it happen?  Where did it happen?  How were people impacted by the heat event?  Which people were most impacted by the heat event?  Students could work in groups of three—where each student reads one of the articles and explains the situation to the other students.    Attachment 1: News stories on heat related deaths  Attachment 2: Question sheet for students on News stories of heat related deaths  In addition to news articles or instead of reading, the following are links to two short news videos that document heat related deaths which can be shared with students in addition to the reading or in place of the reading: 18-Year-Old Among 6 Heat-Related Deaths in Chicago <http://abcnews.go.com/US/video/18-year-heat-related-deaths-chicago-14137199>  French 2003 heat related deaths  <http://abcnews.go.com/Archives/video/aug-17-2003-heat-related-deaths-france-10487403>  After reading the articles and/or watching the video, lead the class in a discussion of the impact of extreme temperatures on human health and specifically on the number of deaths that occur during heat events. Ask the students to summarize what happened in each city. Help students to conclude that when extreme weather events occur, there is an increase in the number of deaths that also occur. Students may also observe that more deaths occur among elderly or sickly individuals and among homeless or people without appropriate shelter and access to cooling centers.  Heat waves can kill!  Write this working definitions on the board:  **Heat wave**: A heat wave is a prolonged period of excessively hot weather. In some areas, heat waves may include periods of high humidity. There is no universal definition of a heat wave--the term refers to the weather relative to the usual weather in an area.  Note: People do adapt—what is considered hot in one place is cool for another. The definition of a heat wave will vary by location. Independent of the location, when it gets hotter than normal for that location, more people tend to die of heat related issues.  **Explore:**  Ask students why a community might care about what happens when it gets hot and stays hot for several days. In particular, why would a community care about heat related deaths? Answers could include: I don’t care—I’m healthy and it won’t impact me or it might impact my elderly relatives, it won’t happen here because it doesn’t get that hot.  Lead the discussion to talk about how temperatures around the world are rising. Let’s look specifically at North America. Why is tracking extreme heat related health issues in North America becoming increasingly important? It is because the temperatures outside across North America are getting warmer. Show students a graph of average temperatures taken over the last 150 years. Note this trend is occurring around the world. You could use other NASA charts to make this point.  NASA has published data for global warming as annual trend charts and as surface temperature anomaly maps. As shown in the chart below, the global warming trend for the past 125 years has been an almost steady rise. During 2005 the highest global surface temperature in those 125 years of record keeping was recorded. NASA comments "Record warmth in 2005 is notable, because global temperature has not received any boost from a tropical El Niño this year. The prior record year, 1998, on the contrary, was lifted 0.2 degrees C above the trend line by the strongest El Niño of the past century. Increasing global surface temperature - particularly temperature in the Gulf and areas of Atlantic Storm formation are significant as they could lead to more powerful hurricanes in the future. Full details of the NASA data can be seen at their [Surface Temperature Analysis website.](http://data.giss.nasa.gov/gistemp/)  Surface temperatures in North America  *global-warming-graph* Graph by NASA  Temperature increase in degrees C  Decades  Help the students conclude: Temperatures in North America and around the world are increasing. If the death rate increases when the temperature increases, then as the temperature rises—even by 1-2 degrees Celsius, communities are going to see more people dying during heat related events—at least in the short term.  **Explain:**  Tell students that the scientists at Johns Hopkins University in Washington D.C. were very interested in understanding the impacts of air pollution and climate on morbidity and mortality in North America. If temperatures were increasing, they wanted to know what that meant to the people who lived there. To figure that out, they needed data.  Write these definitions on the board:  **Climate** is the long term trend in the weather for a specific area.  **Morbidity** is the incidence or prevalence of a disease or of all diseases in a population. Morbidity is often measured by any quantifiable incidence of disease short of death such as hospital admissions or doctors’ office visits for a particular ailment.  **Mortality** is the rate of death.  In a project funded by the Health Effects Institute (HEI), 108 North American cities were studied for 14 years. Data was collected for these cities on a daily basis covering a variety of climate and pollution related items. The data was then analyzed in a number of ways to determine the impacts of air pollution on death rates.  Explain to the students that your job for this lesson will be to help analyze this real life data set to investigate how heat waves and temperature impact death rates. You will be the statisticians responsible for analyzing the data and drawing some conclusions about the graphs you create.  As a class, we will study one city, Chicago, Illinois and analyze the data points on death and temperature for that city.    Here’s what the scientists at Johns Hopkins University did—they collected data for 14 years -- 5114 data points for each city for daily temperature and death rate. That is a lot of data for any statistician to manage. In order to have more accurate data (data that has less variability from one day to the next) and to make the data easier to work with, the data needs to be averaged by groups of 100 data points. The data set we are working with has been sorted by temperature and can be found in Attachment 3. Attachment 3 contains the entire data set of 5114 points for the day, month, year, temperature and death rate for the city of Chicago, Illinois. This will be the data set for all our calculations.  Step 1 in analyzing the data: Calculate the average temperature and death rate for the first 100 points. Using a subset of the data in attachment 3b (the first 100 points), students will add up the temperatures and divide by 100. They will repeat the same step for the death rate. This will give students their first pair of data points. The data set we are using includes 5 items: the date (year, month, and day), the temperature, and the number of deaths that occurred that day. Adding up the first 100 temperatures and dividing by 100 will result in an average of 3.22 degrees C. Adding up the number of deaths per day and dividing by 100 will result in an average of 127.27 deaths.  Ask the students: Why group the data into sets of 100 data points? We started our study of temperature and death rates with 5114 data points—which is a lot of points to graph. In looking at a scatterplot for all of the 5100+ data points, we lose some information because there is so much variation in the individual points. By grouping the points together and taking an average, we reduce some of the day to day variation and therefore reduce the amount of error we have in any given point. Grouping the data by 100 is arbitrary—we could have used 25, 50 or 75 or 200. Grouping by 100 gave us 51 points which is enough to graph in a scatterplot, shows the shape of the data curve, and is not overwhelming. With a more advanced group of students, you could have the students vary the number of points in a group to see what happens to the shape of the curve.  While students could calculate averages for each set of 100 points by hand, this is a lot of painstaking, redundant work. That is why we have computers! The averaged data points for students to graph can be found in Attachment 4.  Step 2 in analyzing the data: create a scatter plot of the data. Students will next plot the 51 data points on graph paper. They will generate a scatter plot that looks like this:  Death Rate by Temp for Chicago  Number of Deaths  Average Temperatures in degrees F  The title of the chart would be Death Rate by Temperature for Chicago. The Y axis is the death rate. The X axis is the temperature. Have students label their scatter plot.  Step three in data analysis: determine a line of best fit for the data. Ask the students to draw a line through the data points that best approximates the data they have plotted on their scatterplot. Results will vary depending on how the students look at the data. Some students will try to draw one line through the data. Other students may see that the death rate goes down until somewhere between the temperatures of 65-75 degrees Fahrenheit. At that point, the death rate increases. Those students may draw two lines. Students will draw many different lines to approximate the slope of the data curve. Discuss the variations and how the students interpret the scatterplot that they drew. The best analysis would be to have two lines—one with a negative slope for the points plotted from 0-70 degrees. The second line would have a positive slope through the points from 70-100 degrees.  Death Rate by Temp for Chicago  **Elaborate:**  Remind students that the purpose of statistics and data analysis is to:   1. Produce trustworthy data 2. Analyze data to make their meaning clear 3. Draw practical conclusions from data.   One tool to help to make sense of data is linear regression. This is a method of fitting a line to data to find the best fit curve. The first step in linear regression is to plot and evaluate a scatter diagram—like the one the students created above. Next, statisticians perform linear regression analysis. This can be done either on a graphing calculator or using Excel. For this lesson, students will use the graphing calculator. The model that is created is then compared to the original data set—if they are similar, then the statistics work is considered complete and the line of best fit is interpreted. If not, modifications are made to the assumptions that go into the linear regression and the analysis is done again.    Is there a temperature or range of temperatures at which the fewest deaths occur? At what temperature or range or temperatures do the most deaths occur? The scatterplot shows a minimum number of deaths in Chicago occurring somewhere between 65-75 degrees Fahrenheit.  Step by step instructions for doing a linear regression using a TI-83/84 programmable calculator can be found in calculator hand book. The steps needed are:   1. Use the statistical list editor to enter the 51 average temperatures into L1 and the 51 average deaths into L2. 2. Plot the data using the STAT Plot function to view a scatterplot of the data. Where is the lowest point of the scatterplot? It should be between 65-75 degrees. 3. Calculate a linear regression.   For the purposes of this lesson, students will do a simple linear regression. More advanced students could split the analysis into two parts—one for the decreasing slope section of the scatterplot (points with temperatures below 70 degrees) and one for the increasing slope section of the scatterplot (points equal to or above 70 degrees).  The graph below shows a more advanced polynomial regression analysis done using Excel. With this analysis, graph shows that the minimum daily deaths occur at about 68 degrees Fahrenheit. Above an average 68 degree temperature, the death rate starts to climb.  The temperature associated with the minimum daily deaths is 68 degrees.  Does this make sense? Yes and no—this is where statisticians must analyze the graphs and data they create. There are many more data points around 70 degrees than there are at the extreme temperatures. When the temperatures in Chicago rise above 90 degrees (and there are a relative low number of those days), the death rate increases at a rapid rate. It does make sense that people would be more stressed at high temperatures and at very low temperatures—and that the death rate, especially among the elderly, young or physically challenged would increase at times of higher physical stress.  **Evaluate:**  For every city, there is a temperature range where the fewest number of deaths occur. If you were in charge of public health, why would this information be important? Is it possible to impact human health as temperatures rise? What kinds of things could you do to minimize the impact of human health on a community?  The City of Chicago cares about this information. Have students read the article from the NY Times that tells about what Chicago is doing to minimize the human impact as temperatures rise:  <http://www.nytimes.com/2011/05/23/science/earth/23adaptation.html?_r=1&hp=&pagewanted=all>  Make a list of at least 5 things people in Chicago are doing to minimize the impact of increases in city temperatures. Add one more idea to the list of something Chicago could do on improve human health issues that will occur as a result of rising temperatures.  **Guided Practice:**  Have students practice their skills using worksheets. Examples of sites with prepared worksheets include:  Averaging:  <http://www.superteacherworksheets.com/mean-averages.html>  Scatterplots:  <http://www.pearsonsuccessnet.com/ebook/workbooks/0-13-037878-X/0-13-122241-4/gpda_index.html>  Linear regression:  <http://www.superteacherworksheets.com/mean-averages.html> |
| **Assessment** | Have the students assigned to groups of 2-3 students. Give the students a second sample of data points for a new city. Data for Atlanta, GA is contained in attachment 5. Columns D and E on the Excel spreadsheet contain the actual temperatures and death rates. Columns J and K have the averaged data. Have them walk through the steps of:   1. Calculating the averages for temperature and death rate for the first the first few terms of the 100 data points. The following are the actually averages—that can used to check the student calculations and for plotting the scatter plot points.  |  |  | | --- | --- | | Average | Average | | Temps | Deaths | |  |  | | 27.98 | 24.98 | | 34.89 | 26.95 | | 37.99 | 25.18 | | 39.91 | 25.76 | | 41.61 | 25.42 | | 43.09 | 24.07 | | 44.46 | 24.89 | | 45.77 | 24.71 | | 46.84 | 24.35 | | 47.82 | 23.83 | | 48.92 | 24.66 | | 49.94 | 24.51 | | 50.95 | 24.29 | | 52.04 | 23.24 | | 53.16 | 23.76 | | 54.27 | 24.28 | | 55.37 | 23.95 | | 56.40 | 23.21 | | 57.43 | 23.62 | | 58.49 | 23.63 | | 59.60 | 23.77 | | 60.64 | 23.81 | | 61.70 | 22.97 | | 62.60 | 22.86 | | 63.37 | 23.60 | | 64.29 | 23.61 | | 65.15 | 23.15 | | 66.11 | 23.11 | | 67.00 | 22.65 | | 68.03 | 22.15 | | 69.08 | 21.83 | | 70.11 | 20.48 | | 71.09 | 22.51 | | 72.04 | 22.61 | | 73.05 | 21.87 | | 73.84 | 22.02 | | 74.73 | 22.65 | | 75.61 | 21.51 | | 76.36 | 21.45 | | 77.09 | 21.45 | | 77.75 | 21.36 | | 78.19 | 21.19 | | 78.81 | 20.96 | | 79.24 | 20.87 | | 79.84 | 21.99 | | 80.43 | 21.98 | | 81.13 | 21.26 | | 81.92 | 22.40 | | 82.81 | 21.86 | | 83.83 | 21.94 | | 85.63 | 22.16 |  1. Creating a scatterplot of the averaged data points 2. Draw a line(s) of best fit on the scatterplot.   There is a slight uptick in the data above 80 degrees Farenheite.   1. Enter the averaged points into the graphing calculator and do a linear regression analysis. An example of a more advanced linear regression done using Excel is shown below. |
| **Critical Vocabulary** | Climate: The pattern of temperature and precipitation typical of an area over a long time period.  Death rate: The ratio of total deaths to total population in a specified community or area over a specified period of time.  Heat wave: A heat wave is a prolonged period of excessively hot weather. In some areas, it may include periods of high humidity. There is no universal definition of a heat wave--the term refers to the weather relative to the usual weather in an area.  Linear regression: The relation between variables when the regression equation is linear: e.g., y = ax + b.  Line of best fit: A straight line drawn through the center of a group of data points plotted on a scatter plot.  Morbidity: The incidence or prevalence of a disease or of all diseases in a population. Morbidity is often measured by any quantifiable incidence of disease short of death such as hospital admissions or doctors’ office visits for a particular ailment.  Statistics: The mathematics of the collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling.  Weather: The state of the atmosphere at a given time and place, with respect to variables such as temperature, moisture, wind velocity, and barometric pressure.  Heat wave: A heat wave is a prolonged period of excessively hot weather. In some areas, it may include periods of high humidity. There is no universal definition of a heat wave--the term refers to the weather relative to the usual weather in an area. |
| **Modifications** | 1. In the engage section of the lesson, have students watch the videos and discuss the problem of extreme temperature and death rates instead of having them fill in the graphic organizer. ELL could be given sentence starters for the graphic organizer. 2. Have students work in groups of 2-3 to do the calculations. This allows students to check their work. Another option is to have students do individual calculations and then share their results with a partner. 3. Simplify the calculations to do a group of 25 or 50 terms for the averaging instead of 100 terms. Another option is to have students do the averaging using Excel. 4. Make the scatter plot as a class. Assign each student 2-3 points to plot on a class graph rather than having students plot the points on their own graph. 5. Use a think-pair-share for the discussions. 6. Link a graphing calculator attached to the class projector and walk the students through each step of the exercise as they do the activity at their desks. 7. As an extension, advanced students can explore other health related statistics. Students can explore some of the types of data collected by looking at this web site <http://www.ncdhhs.gov/factsandfigures/index.htm> . Student can also brainstorm or list the types of data that could be collected by the health department and how they might be used to manage health in the community. 8. Students wanting to explore the impacts of weather related health issues in more depth could look at a city such as New Orleans—and list other weather related issues due to increasing temperatures. <http://geology.com/articles/rebuilding-new-orleans.shtml> |
| **Alternative Assessments** | Students could research and report on how statisticians collect and use other health related data.  Students could create their own “real world” data by creating a school survey. They could create a scatter plot of the data that they collect. |
| **References** | National Morbidity and Mortality Air Pollution Study  [**http://www.ihapss.jhsph.edu/publications/Results/update.main.htm**](http://www.ihapss.jhsph.edu/publications/Results/update.main.htm) |
| **Supplemental Information** | What is a linear regression? Linear regression is a widely used statistical technique that looks at the relationships between two variables, X and Y. For each subject or study, you know both the X and the Y and you want to find the best line through the data. In some situations, the slope and the intercepts have a meaning. In other cases, the line helps to find new values of X given Y or Y given X. In general, the goal of linear regression is to find the line that best predicts Y from X. Linear regression does this by finding the line that minimizes the sum of the squares of the vertical distances of the points from the line.  Linear regression does make some assumptions—in particular that there is a relationship between the data (which can be seen by first looking at a scatterplot of data) and that errors in the data are normally distributed—the data values tend to be about the same distance above and below a mean or average.  The term "regression", like many statistical terms, is used in statistics differently than it is used in other situations or areas of study. The method was first used to examine the relationship between the heights of fathers and sons. The two heights were related, but the slope is less than 1.0. A tall father tended to have sons that were shorter than he was; a short father tended to have sons taller than he was. The height of sons “regressed to the mean” or average. The term "regression" is now used for many sorts of curve fitting.  Note that linear regression does not *test* whether data are linear. It assumes that the data are linear, and finds the slope and intercept that make a straight line best fit your data!  How do we use the linear regression? The goal of linear regression is to adjust the values of slope and intercept to find the line that best predicts Y from X.  How does the population size impact the error in the data? As population size increases, the error in the data goes down.  Why did we group the data into sets of 100 data points? We started our study of temperature and death rates with over 5100 data points—which is a lot of points to graph. In looking at a scatterplot for all of the 5100+ data points, we lose some information because there is so much variation in the individual points. By grouping the points together and taking an average, we reduce some of the variation and therefore reduce the amount of error we have in each point. Grouping the data by 100 is arbitrary—we could have used 50 or 75 or 200. Grouping by 100 gave us 51 points which is enough to graph, shows the shape of the data curve, and is not overwhelming. With a more advanced group of students, you could have the students vary the number of points in a group to see what happens to the shape of the curve.  What other health related factors influence the shape of the curve? When epidemiologists at Johns Hopkins University and other universities studied this data, they did consider more factors that could have impacted the shape of the curve—and they made adjustments for those factors to the data and to their conclusions. For example, more deaths are expected during flu season. Since flu season tends to occur during the winter when the temperatures are cold, we would expect to see a higher death rate during times of colder temperatures. Another factor that epidemiologists consider when looking at this kind of data is the seasonality—no one is certain why, but more people tend to die on Monday that on any other day of the week. As an extension to this lesson, students could brainstorm other variables that might impact the shape of these curves. |
| **Comments** | The purpose of this lesson was to provide an example of using real world data to do statistical analysis. The focus of my Kenan Fellowship was to find ways to bring more advanced math examples into the classroom for math and science classes. This lesson could be used as a part of an interdisciplinary lesson between science and math with a science unit about climate change. |
| **Author Info** | The author of this lesson is Lori Craven, an 8th Grade Science Teacher at Cedar Creek Middle School, Franklin County, North Carolina. |