Teacher Resources

U.S. Nuclear Stockpiles during the Cold War – A Regression Exercise

Developed By Richard Gill, Peabody Veteran's Memorial High School, Peabody, Massachusetts

2011 Naval Historical Foundation STEM Teacher Fellowship

Information: This lesson was designed for AP Statistics students - but it could be used – with minor modifications – for any high school statistics course that covers the topic of regression. The lesson as written asks student to re-express a data set to make it more linear and allow for regression analysis, but this section of the lesson could be removed for non-AP students.

Massachusetts State Standards:

(AI = Algebra 1, AII = Algebra 2, G = Geometry, PC = Pre-calculus)

AI.P.2 – Use properties of the real number system to judge the validity of equations and inequalities, to prove or disprove statements, and to justify every step in a sequential argument.

AI.P.3 – Demonstrate an understanding of relations and functions. Identify the domain, range, dependent, and independent variables of functions.

AI.P.4 - Translate between different representations of functions and relations: graphs, equations, point sets, and tabular.

AI.P.5 - Demonstrate an understanding of the relationship between various representations of a line. Determine a line's slope and x- and y-intercepts from its graph or from a linear equation that represents the line. Find a linear equation describing a line from a graph or a geometric description of the line, e.g., by using the "point-slope" or "slope y-intercept" formulas. Explain the significance of a positive, negative, zero, or undefined slope.

AI.P.11 - Solve everyday problems that can be modeled using linear, reciprocal, quadratic, or exponential functions. Apply appropriate tabular, graphical, or symbolic methods to the solution. Include compound interest, and direct and inverse variation problems. Use technology when appropriate.

AI.D.1 - Select, create, and interpret an appropriate graphical representation (e.g., scatterplot, table, stem-and-leaf plots, circle graph, line graph, and line plot) for a set of data and use appropriate statistics (e.g., mean, median, range, and mode) to communicate information about the data. Use these notions to compare different sets of data.

AI.D.2 - Approximate a line of best fit (trend line) given a set of data (e.g., scatterplot). Use technology when appropriate.

AII.P.5 – Perform operations on functions, including composition. Find inverses of functions.

AII.P.6 – Given algebraic, numeric and/or graphical representations, recognize functions as polynomial, rational, logarithmic, or exponential.

AII.D.1 – Select an appropriate graphical representation for a set of data and use appropriate statistics (e.g., quartile or percentile distribution) to communicate information about the data. (12.D.2)

G.G.11 – Demonstrate an understanding of the relationship between various representations of a line. Determine a line's slope and x- and y-intercepts from its graph or from a linear equation that represents the line. Find a linear equation describing a line from a graph or a geometric description of the line, e.g., by using the "point-slope" or "slope y-intercept" formulas. Explain the significance of a positive, negative, zero, or undefined slope.

G.G.12 – Using rectangular coordinates, calculate midpoints of segments, slopes of lines and segments, and distances between two points, and apply the results to the solutions of problems.

G.G.13 – Find linear equations that represent lines either perpendicular or parallel to a given line and through a point, e.g., by using the "point-slope" form of the equation.

PC.D.2 – Apply regression results and curve fitting to make predictions from data.

Common Core Standards:

S-ID.6 - Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

- Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.
- Informally assess the fit of a function by plotting and analyzing residuals.
- Fit a linear function for a scatter plot that suggests a linear association.

S-ID.7 - Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

S-ID.8 - Compute (using technology) and interpret the correlation coefficient of a linear fit.

S-ID.9 - Distinguish between correlation and causation.

Lesson Plan:

Following WWII and the expansion of nuclear weaponry, the U.S. began to stockpile warheads in an effort to "stay ahead" of the Soviet Union's capabilities. These military buildups constituted the Cold War which ran from 1946 to 1991.

For more information about nuclear weapons during the cold war, click HERE.

The table below shows the number of nuclear warheads in the U.S. arsenal from 1945 to 2012; some data is missing because it is classified, the data in the later years are estimates.

Date	# of U.S. nuclear warheads
1945	6
1946	11
1947	32
1948	110
1949	435
1950	369
1951	640
1952	1005
1953	1436
1954	2063
1955	3057
1956	4618
1957	6444
1958	9822
1959	15468
1960	20434
1961	24156
1962	27305
1963	29049
1966	32193
2002	10600
2009	2200
2012	1900

Part I: Using technology (statistics software or your calculator) create a scatter plot of the data and describe the relationship between the variables ***year*** and ***number of warheads***.

Part II: Identify any outliers (visually) and remove them from your data set. Using technology, run a regression on the remaining data points. Discuss the validity of your regression to predict the number of warheads per year. What are your concerns? Will your model make accurate predictions? Is your model "useful" at all.

Part III: Re-express the data (leave the outliers out!) to find the best linear model possible to explain the relationship and make predictions.

- Write the equation of the model and find the number of warheads in 2002 ***as predicted by the model***.
- Discuss the validity of your model to predict the number of warheads per year; be sure to address r^2.
- Interpret the slope and the intercept of your linear model in context.
- What are your concerns about your model? Do you believe your model will make accurate predictions? Is your model "useful" at all?

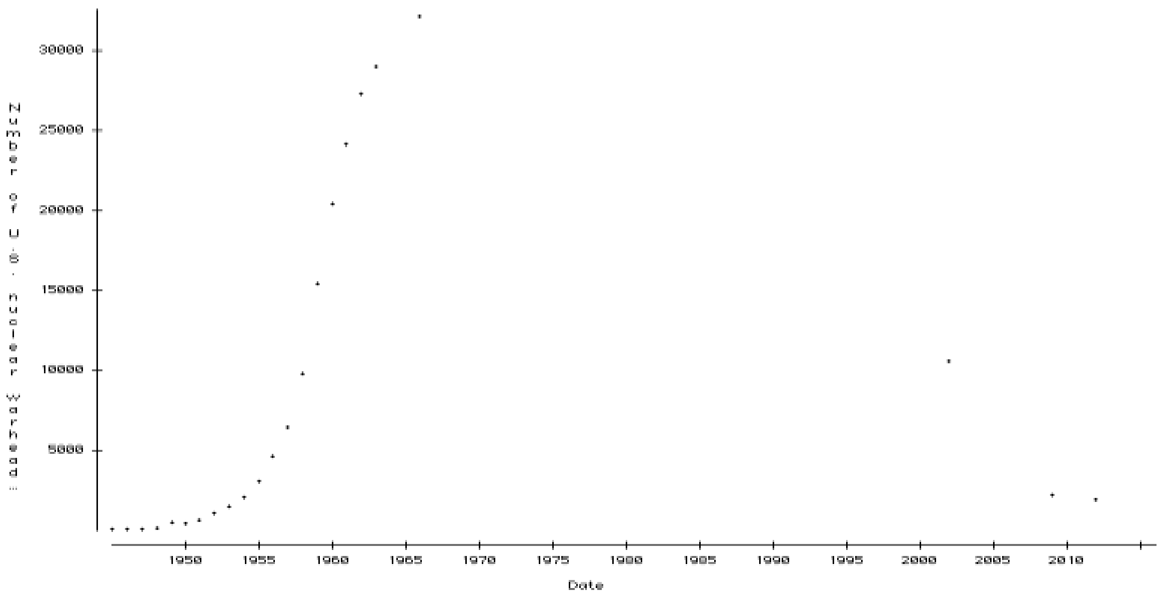
Part IV: Clearly the U.S. never had as many warheads as your model predicted in Part III. In fact, the original data set contained the actual number of U.S. warheads in 2002. Explain, using your understanding of U.S. History, what the data set probably looks like through the years of missing data.

Answers / Teacher's Resources

This statistics exercise is designed for students to complete on their own or in small groups. Students should have access to a computer with statistical software or calculator with statistics capabilities. The exercise serves several purposes. For AP Statistics, it gives students the opportunity to work on their regression and re-expression skills on real-world data that is clearly non-linear. It then asks students to look at the data a second time in the context of real-world history (the Cold War) and evaluate the data in context of the history. This critical thinking about the data is very important and difficult for high school students. The author's expectation is that most students will, initially, "jump right in" the re-expression process without really THINKING about the data they've been given. Part IV of this lab forces students to look at the data a second time and make critical, real-world connections to American History.

NOTE: Additional computations / graphs / analysis can be found in the accompanying, multi-tab MS Excel spreadsheet file called "Re-expression Lab Answers and Graphs."

Part I: Using technology (statistics software or your calculator) create a scatter plot of the data and describe the relationship between year and number of warheads.



Direction: positive
Strength: moderate to strong
Form: non-linear

Part II: Identify any outliers (visually) and remove them from your data set. Using technology, run a regression on the remaining data points. Discuss the validity of your regression to predict the number of warheads per year. What are your concerns? Will your model make accurate predictions? Is your model "useful" at all.

The points 2002, 2009 & 2012 appear to be outliers in the x-direction and they have been removed as instructed. Regression analysis was completed, but the data is clearly non-linear (this is confirmed by the residual graphs as well). The model cannot be used to make accurate predictions because the linearity assumption is violated.

Part III: Re-express the data (leave the outliers out!) to find the best linear model possible to explain the relationship and make predictions.

- Write the equation of the model and find the number of warheads in 2002 ***as predicted by the model***.

Equation: $log(\hat{y}) = -700.637 + 0.361735(year)$
Solution when x = 2002:
$log(\hat{y}) = -700.637 + 0.361735(2002)$
$log(\hat{y}) = 23.55647$
!!!!!!!!!!!!!!!!! $\hat{y} = 3.6x10^{23}$!!!!!!!!!!!!!!!!

- Discuss the validity of your model to predict the number of warheads per year; be sure to address r^2.

The re-expressed data is much more linear; but it may not be "straight enough" based on significant curvature / non-randomness of the residual plots. That said, I would be much more comfortable using the re-expressed model to predict the number of warheads, especially between 1948 and 1960 where the graph appears most linear. It might be best to consider two separate regressions based on the original scatterplot's curvature…but that's beyond the scope of this class! The r-squared value is over 90%, so if we can say the data is str8-enough, more than 90% of the variability in the number of wareheads can be explained by the progression of years.

- Interpret the slope and the intercept of your linear model in context.

Slope = 0.361735. For each year's increase, we expect that the Logarithm of the number of warheads will increase by 0.361735 (minimally-acceptable answer).

Our model predicts that each year the U.S. nuclear arsenal increased by between two and three warheads (better answer which requires solving the log!).

Intercept = -700.637. In year ZERO, the US nuclear stockpile had negative 700 warheads; in context this number is meaningless; it merely gives us a way to get into the scale of the data we're examining.

- What are your concerns about your model? Do you believe your model will make accurate predictions? Is your model "useful" at all?

By examining the scatter plot, the data appear MORE linear than before, but the post-regression residuals still show significant curvature (non-randomness); cannot be certain that the linearity assumption has been met. That said, our model should do a decent job of estimating # of warheads during the Cold War period.

Part IV: Clearly the U.S. never had as many warheads as your model predicted in Part III. In fact, the original data set contained the actual number of U.S. warheads in 2002. Explain, using your understanding of U.S. History, what the data set probably looks like through the years of missing data.

The data that's missing is from 1966 to 2000 or so; this set of dates corresponds roughly to the Cold War period of U.S. history where the two world super-powers (the USA and USSR) were engaged in an arms race and "M.A.D." (mutually-assured destruction) policies. Both nations were building their nuclear arsenal in an attempt to gain an advantage should war ever erupt. In the end, the economy of the USSR was unable to sustain the nearly exponential growth of their military and their economy crumbled, bringing an end to the Cold War under President Reagan. In all likelyhood, the data that's missing (probably due to security concerns during the post-Cold War era) continues in a downward, curved manner so the entire data set looks like a parabola. The data points eliminated early in this analysis are probably not outliers! The decreases likely are a result of the post-Cold War era arms reduction treaties between the US and USSR/Russia like SALT.

Additional Historical Background/Perspective/Extension Discussions:

Some argue that the massive expenditures during the Cold War were an absurd waste of national resources; others that these expenditures were necessary to prevent nuclear holocaust and protect democracy – priceless outcomes so the cost does not matter. The truth probably lies somewhere between these two arguments; but there can be no doubt that Cold War expenditures into the nuclear program helped spur development in technologies and industries that would not otherwise exist today.

- Weapons targeting required precise positioning of mobile (submarine-based) ballistic missiles, so the military developed and launched satellite positioning; today's GPS.

For more information about navigation, click HERE.

- The need for stealth, self-sufficiency, and the ability to patrol independently for long periods led to the perfection of safe nuclear plants that were the predecessors of atomic energy generation plants in use in the U.S. and around the world.

For more information about nuclear propulsion, click HERE.

- Accurate, reliable and secure communications were crucial to ensure that the President could control launch orders for deployed nuclear weapons. This requirement led to improvements in acoustics, equipment and satellite technology leading to the advent of the cell phone and satellite radio and television. Furthermore, control of complicated vessels, long-range nuclear weapons, nuclear plants and satellites necessitated significant advancements in computer technology.

For more information about command, control and communications, click HERE.

- Fear of the enemy's ability to destroy communications on a large scale by attacking a single, important computer station lead to development of the internet – accessible from various locations with the proper equipment.

For more information about the creation of the internet, click HERE.

Without the Cold War – and the U.S. military spending during the Cold War - many of today's modern conveniences would not exist.

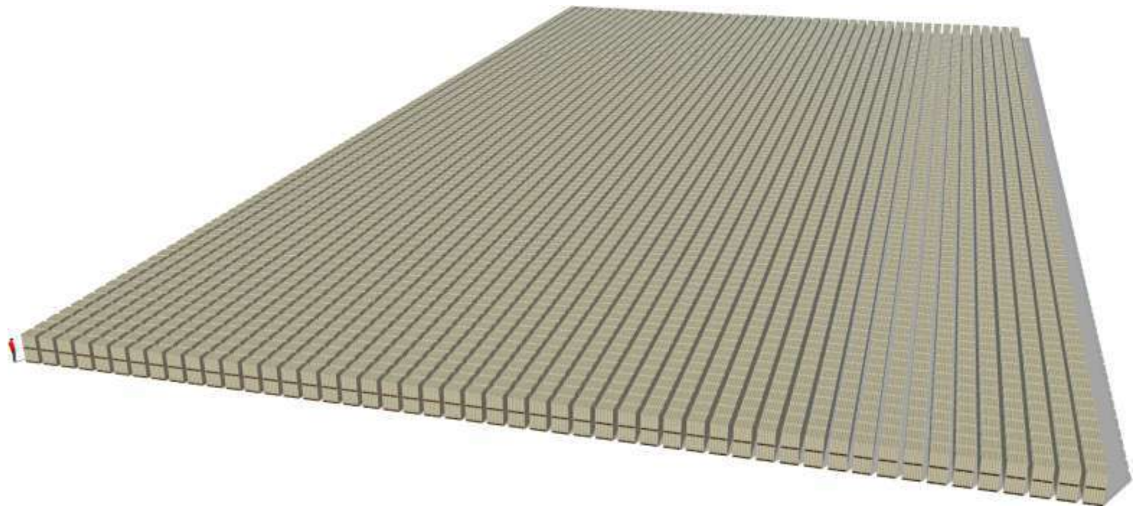
The amount of money spent by the U.S. government on "defense" during the cold war is difficult to pin down, but can be roughly estimated at in excess of 22 TRILLION DOLLARS ($22,000,000,000,000). You can use the following diagrams to help students understand how much money that is!

Ten Thousand Dollars

One Million Dollars

One Billion Dollars

That's the man!!



One Trillion Dollars

Now imagine 22 of these!