Making **G.A.I.N**.s (**G**lobal **A**wareness by **I**nvestigating **N**umbers)

Algebra 1 – Using Linear Models to Investigate Syrian Refugee Crisis (Part 2)

|  |
| --- |
| ***ORIGINAL Guiding Question:*** *How can using a linear model from data collected in 2015 help Turkey predict the number of refugees from Syria it can expect to support by the end of 2017?* |

|  |
| --- |
| Last time we examined the following data regarding movement of Syrian refugees into Turkey from January, 2015 through November, 2015. We used two of the given points to determine a line of best fit and used that line to calculate the two missing values from the table. We also used the line to predict the number of refuges Syria could expect by the end of 2017.  |
|  |

|  |
| --- |
| ***NEW Guiding Questions:*** * *After calculating the LSRL, does the correlation coefficient confirm that a linear regression is the best model?*
* *What does a correlation coefficient tell you about the strength of a linear regression line?*
* *How does an increase in data affect the validating of the prediction? (Extrapolation)*
 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. Enter the data from the table below into the graphing calculator.
2. Enter Time into $L\_{1}$, using Time 0 as January, 2012, and every input as the number of months since January, 2012. (For example, February, 2013 would have an input of $x=13$, as it is 13 months past January, 2012.)
3. Enter the number of refugees as a value in 1000’s (For example, 456,874 refugees would have an output value of 456.874.)

|  |  |
| --- | --- |
| Time in Months **(Months since January, 2012)** | Number of refugees from Syria entering Turkey |
| January, 2012 **(0)** | 9,500 |
| August, 2012 **(7)** | 80, 410 |
| January, 2013 **(12)** | 150, 906 |
| July, 2013 **(18)** | 412, 789 |
| January, 2014 **(24)** | 569, 199 |
| September, 2014 **(32)** | 847, 266 |
| November, 2014 **(34)** | 1,165,279 |
| March, 2015 **(38)** | 1,698, 472 |

 |
| 1. Turn the Stat Plot on and alter the window to see all given data points.
2. Describe the association.
 |
| 1. Turn on diagnostics and run a Linear Regression in order to determine an LSRL, a correlation coefficient, and the $r^{2}$ value. Type the equation into $“y="$ and check the graph to make sure the LSRL is correct.

LSRL:r:$r^{2}$:***Check with your group to make sure that you all obtained the same slope and y-intercept to verify you all typed in the correct information.***1. What does the slope mean in context of the situation?
2. State the y-intercept as a coordinate and explain what it means in context. Does the y-intercept make sense? Explain.

1. Calculate the residuals in $L\_{3}$ and then go back to the stat plot and change $L\_{2}$ to $L\_{3}$ in order graph a residual plot. Alter the window to see all data points. **Make a sketch of your residual plot.** Based on the residual plot, is it reasonable to use a linear regression to model the data? Explain.
2. What does the correlation coefficient tell you about the linear model?
 |
| 1. Given that the number of refugees in March, 2018, was 3,547,194, calculate the residual for that month. Explain the significance of the value in context of the situation.

  |
| 1. In addition to the data given earlier, your team gathered additional data by the UNHCR (United Nations Refugee Agency). Add this new data to your table following the same structure as before with respect to input and output values.

|  |  |
| --- | --- |
| Time in Months **(Months since January, 2012)** | Number of refugees from Syria entering Turkey |
| October, 2015 **(45)** | 2, 072, 290 |
| January, 2016 **(48)** | 2, 503, 549 |
| August, 2016 **(55)** | 2, 726, 980 |
| January, 2017 **(60)** | 2, 854, 968 |
| June, 2017 **(65)** | 3, 049, 879 |
| January, 2018 **(72)** | 3, 466, 268 |
| March, 2018 **(74)** | 3, 547, 194 |

 |

|  |
| --- |
| 1. Run a new Linear Regression in order to determine an LSRL, a correlation coefficient, and the $r^{2}$ value. Type the equation into $“y="$ and check the graph to make sure the LSRL is correct.

LSRL:r:$r^{2}$:***Check with your group to make sure that you all obtained the same slope and y-intercept to verify you all typed in the correct information.***Since you have a NEW LSRL, calculate NEW residuals in $L\_{3}$. Create a new residual graph. Alter the window to see all data points. Based on the NEW residual plot, is it reasonable to use a linear regression to model the data? Explain. After seeing this residual plot, are you more comfortable using a linear model now than previously in question 3C? What does the correlation coefficient tell you about the linear model? |
| 1. Calculate the new residual for March, 2018. Is this value different than what you found previously question 4? Why? Which prediction gave the lower residual (a more accurate prediction)? Why accounts for the difference?
 |
| EXTENSION:Experiment with a quadratic or exponential regression to model the data and use the residual plot to visualize how well those regressions model the data. |

Introduction to GapMinder

1. Go to [www.gapminder.org/world](http://www.gapminder.org/world)
2. What do you notice?
3. What do the colors on the graph represent?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Red | Lt. Blue | Dk. Blue | Yellow | Green | Orange |
|  |  |  |  |  |  |

1. What information do you gain by putting your mouse over a colored circle on the graph?
2. What is the independent variable ($x)$?
3. What does the dependent variable ($y)$ represent?
4. Push the PLAY button. What is this graphical representation telling you?
5. Select United States on the right side of the graph. After pushing the play button, you can gather information from the graph by hovering over a point. The data will show up in a white box on the x and y axis. Select five data points and record them below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Income |  |  |  |  |  |
| Life Expectancy |  |  |  |  |  |
| Year |  |  |  |  |  |

1. What was the highest GDP in 2011? \_\_\_\_\_\_\_\_\_\_\_\_\_ Which country? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. What was the highest life expectancy in 2011? \_\_\_\_\_\_\_\_\_\_ Which country? \_\_\_\_\_\_\_\_\_\_\_\_
3. As the GDP of a country increases, what tends to happen with life expectancy? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. Change the variables on the x- and y-axis to a different topic. Select a country that interests you and hit play. Fill in the table to represent your data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

1. What is your independent variable ($x$)?
2. What is your dependent variable ($y$)?
3. Give one interesting observation regarding your new data.