Lizzy Compton, Jacob Koors

Dr. Long

MAT 375

01/26/18

Mini-Project #1

Group 2, Tabligbo

1. Determine if either of your maximum and minimum temperature time-series demonstrate significant increases in temperature over time. Discuss any characteristics of the data that seem relevant.
   1. Both the minimum and maximum annual temperature time series for Tabligbo demonstrate significant increases in temperature over time. This can be seen by running a basic OLS regression in Excel where are results are as follows:
      1. Annual Minimum Temperature Results:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | df | SS | MS | F | P-Value |
| Regression | 1 | 8.758998285 | 8.758998285 | 67.76224818 | 4.77949E-11 |
| Residual | 53 | 6.850819174 | 0.129260739 |  |  |
| Total | 54 | 15.60981746 |  |  |  |

At any reasonable significance level, there is sufficient evidence to conclude that there is a significant relationship between year and the annual minimum surface temperatures in Tabligbo. This relationship is a positive linear relationship, indicating that as year increases, the annual minimum temperatures also increase.

* + 1. Annual Maximum Temperature Results:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | df | SS | MS | F | P-Value |
| Regression | 1 | 12.14939613 | 12.14939613 | 78.87161238 | 4.50309E-12 |
| Residual | 53 | 8.164128708 | 0.154040164 |  |  |
| Total | 54 | 20.31352484 |  |  |  |

At any reasonable significance level, there is sufficient evidence to conclude that there is a significant relationship between year and the annual maximum surface temperatures in Tabligbo. This relationship is a positive linear relationship, indicating that as year increases, the annual maximum temperatures also increase.

|  |  |  |
| --- | --- | --- |
| **Descriptive Statistics** | **Minimum** | **Maximum** |
| Mean | 22.54744984 | 32.84878039 |
| Median | 22.72322389 | 32.84352023 |
| Range | 2.348682796 | 2.576218638 |
| Standard deviation | 0.537652949 | 0.613332199 |
| Count | 55 | 55 |

From these statistics we see that the minimum and maximum datasets have means very similar to their medians, indicating that the data in both has very little skewness either way. It is also important to note that the count for each data set is the same, meaning that we have a consistent number of observations across both datasets. One last thing to note is that the ranges from the two datasets are similar in size to each other, indicating that the datasets are both changing at a similar rate. This is further displayed in our models of the data.

1. Provide and discuss your best models Temperature(time) for each of the two time-series.
   1. The best model for Temperature(time) for the minimum time series is the linear model represented by the equation:
      1. Temperature=21.8696 + 0.025(time)

This model further shows that there is a significant increase in the annual minimum temperatures as the confidence interval for the linear term is both significant and positive. With 95% confidence, we estimate that for each additional year, the annual minimum surface temperatures in Tabligbo increases by between .019 and .031 degrees Celsius.

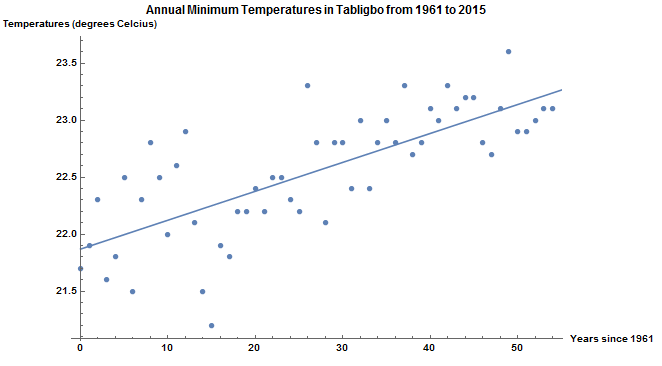
To be sure that this was the best model, we tested for a quadratic term but when we added in a quadratic element we see that this term is not significant. Thus the linear model is the best model.

* 1. The best model for Temperature(time) for the time-series of the maximum temperatures is the linear model represented by the equation:
     1. Temperature=32.04+.0296(time)

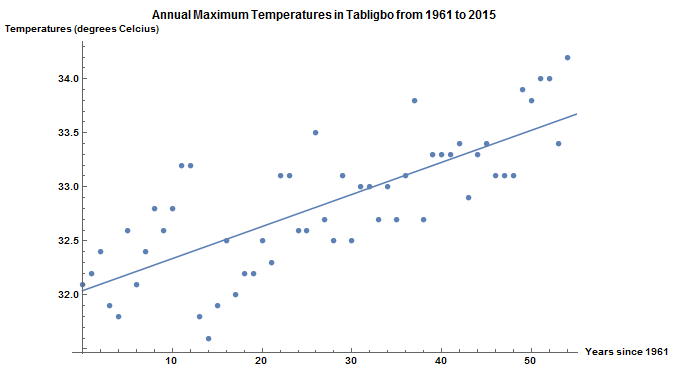
This model further shows that there is a significant increase in the annual maximum temperatures as the confidence interval for the linear term is both significant and positive. With 95% confidence, we estimate that for each additional year, the annual maximum surface temperatures in Tabligbo increases by between .023 and .036 degrees Celsius.

Again, we tested for a significance in a higher order model, but again when we add in a quadratic element we see that the term is not significant. Thus the linear model is also the best model.

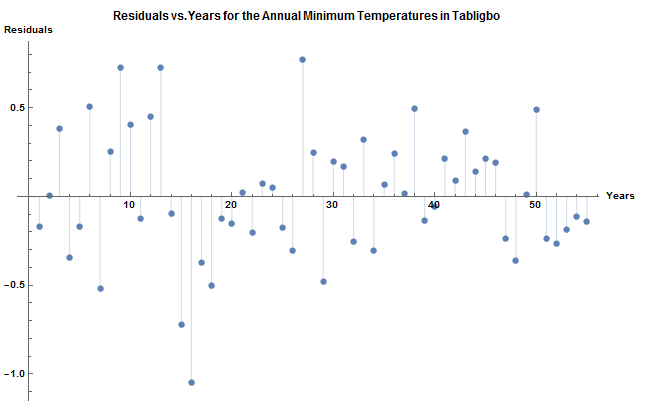
1. Provide graphs of the data with their model(s), with labels and title.
   1. Annual Minimum Temperatures in Tabligbo
      1. Temperature=21.8696 + 0.025(time)

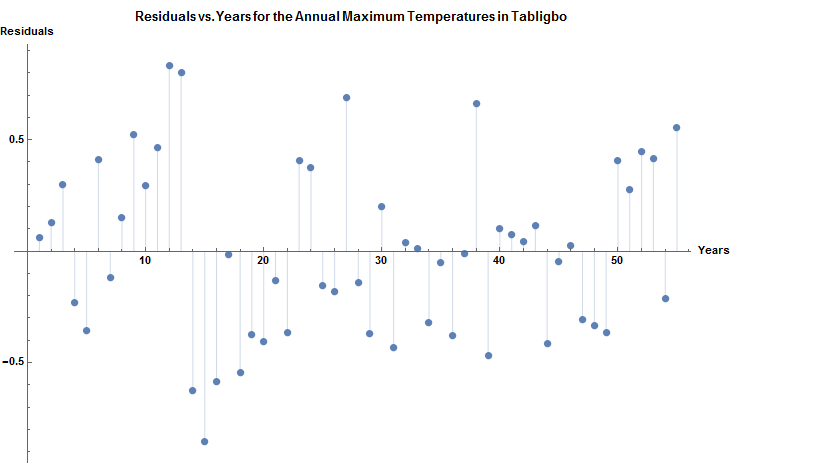


* 1. Annual Maximum Temperatures in Tabligbo
     1. Temperature=32.04+.0296(time)



1. Provide graphs of the residuals, with labels and title.





1. Discuss what additional information might be useful moving forward, problems with your data, and any information you'd like to know from the Togolese meteorologists.
   1. Looking at the graph of our model, we noticed that very few points of our data lie close to our line of best fit. Because of this, we feel that a more complicated model might be able to better explain the variation among our points. As a result, we feel that more analyses should be performed to determine what this model could be. While we have ruled out a quadratic model for both sets of data, there are still more models that we have yet to test that could better fit our data.
   2. Due to the amount of variation in the data, it would be helpful to know how the data was collected so that we can ensure that this variation is due to the data itself and not human error. Is it collected once a day, or once a week, or even once month? For each collection of data, is it collected around the same time of day (in the morning, or at noon, or in the evening), consistently? All of these things affect the data as well as affect how we perform analyses on the data. Also, the geography of where the samples are taken would need to be consistent for accurate and proper data collection. Even within a single city, there is quite a bit of variation and taking a sample near the infrastructural hub of a city versus taking a sample near a large body of water, for example, would greatly skew the results. It would be important to hold all of the variables constant, so that an accurate representation of temperatures can be gathered. If these factors are not consistent we need to know this so that we can adjust our calculations accordingly.