Maria McMahon and Jacob Englert Mini Project 1

Dr. Long

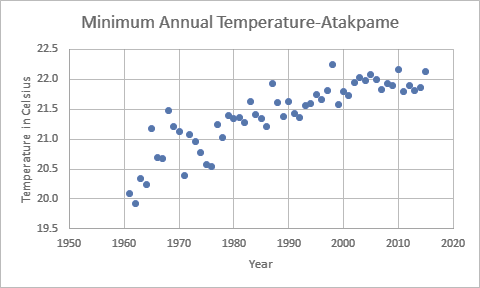
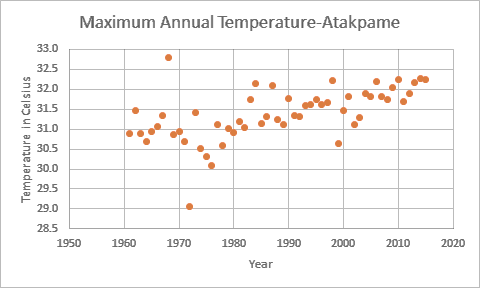
Mat 375

26 January 2018

The country of Togo is located in West Africa on the Gulf of Guinea with a population of nearly 8 million people. The majority of Togo’s economy is dependent on agriculture, so naturally studying the change in climate in the area is crucial not only to the nation itself, but also to those who trade with Togo. The Togolese government is interested in studying a particular phenomenon - the rise in temperature over time. We have been given yearly data going back to 1961 that describes the average annual minimum and maximum temperatures for the city of Atakpamé, the fifth most populous city in the country. Our task is to determine if either the average annual minimum or average annual maximum temperatures have shown significant increase since 1961. To do this, we will use a software package called Minitab to conduct regression analyses on both sets of data separately.

*Atakpamé is located near the center of Togo.*

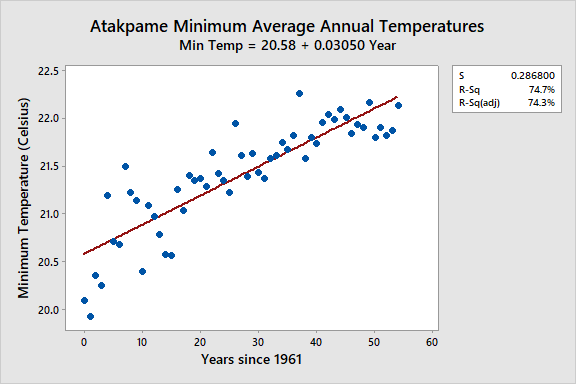
Firstly, let's examine scatterplots of the temperatures over time.



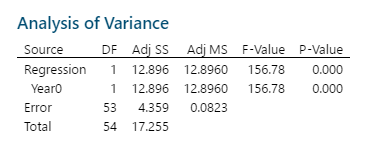
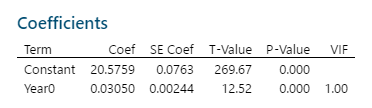
The scatterplot of maximum temperatures appears to show a steady increase over time. There is a moderately strong positive relationship between time (years) and maximum temperature in degrees Celsius, implying that as time goes on, maximum temperatures increase. We do see what may appear to be outliers in 1968 and 1972. We believe a simple linear trend model will be appropriate for fitting this data.

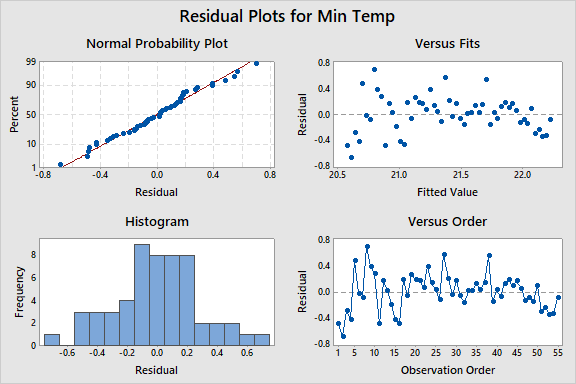
The scatterplot of minimum temperatures appears to show a moderately strong positive correlation between time (years) and temperature in degrees Celsius, implying that as time goes on, minimum temperature rises. However, the increase does seem to be diminishing, leading us to believe that fitting a quadratic model may be a better option than a simple linear model. There don’t appear to be any outliers or discontinuities.

The following are the results of fitting a simple linear regression "best-fit" line to the minimum temperatures data:



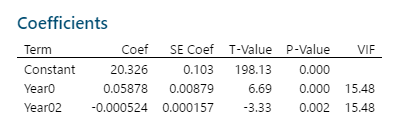
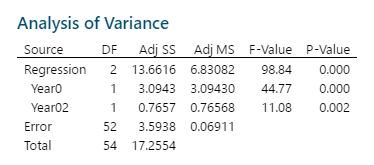
Regression analysis of minimum temperature against time resulted in a significantly non-zero positive slope coefficient (p-value < 0.000), meaning that minimum temperatures are indeed rising with time. This model shows that 74.7% of the variation in the temperature can be explained by the relationship to time. The residual plots can be found on the following page.



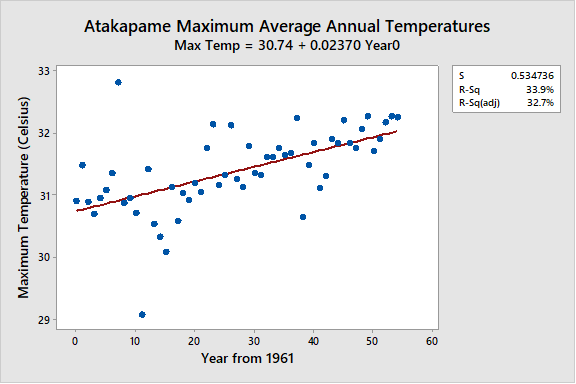
 From the normal probability plot (top left), we see that the residuals likely follow a normal distribution. However, the Versus Fits plot (top right) shows an arc pattern. Originally, we believed that modeling this data would require a model with curvature, and this one obviously does not. Therefore, we will try fitting a quadratic model to the data next. The results can be seen below:

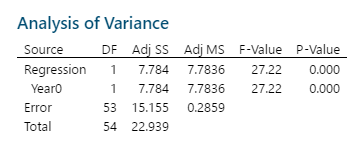
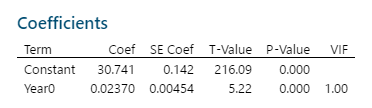


The quadratic model turns out to be significant. This model outperforms the previous as it has successfully reduced the sum of the squared errors through the addition of a significantly non-zero parameter (the squared term). The p-value for this term was very small, which implies that this term added significant information that the first model failed to capture. This model shows 79.2% of the variation in temperature is due to time, compared to the 74.7% from the first model. Residual plots for the model are shown on the next page.

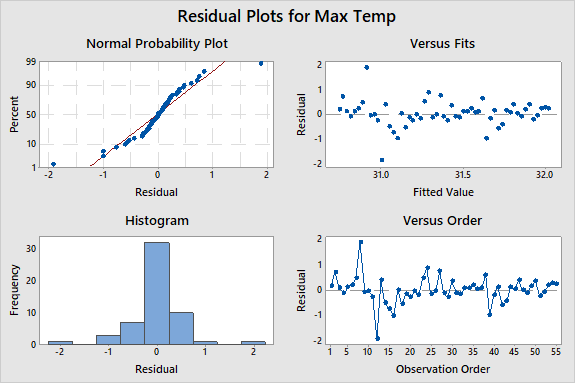


The Normal Probability Plot shows that the assumption of normality is likely fair. Non-constant variance may still be an issue, but we no longer see the need for curvature to be added to the model.

The following results were obtained by fitting a simple linear regression model to the maximum temperature data:



The results of the regression analysis of maximum temperature versus time indicate that maximum temperatures are significantly rising as years pass. Only 33.9% of variation can be attributed to the progression of time, this implies there are other factors affecting the rise in temperature. The residuals plots can be found on the next page.

From the Normal Probability Plot (top left) we can see the residuals mostly follow a normal distribution, with the exception of the suspected outliers causing disruption. The Versus Fits plot (top right) also suggests constant variance in the residuals, with the exception of the two suspected outliers. Overall the simple linear model was a good predictor for the maximum temperature data as the results were shown to be significant.

In conclusion, we have detected significant increases in both annual average minimum and annual average maximum temperatures over the course of the last 55 years. We found a quadratic model (predicting temperature in degrees Celsius with time as the explanatory variable) to be the best fit for the minimum temperature data, and the regression using this model was able to explain 79.2% of the variation in the data. This model fit well, however non-constant variance in the residuals may be an issue to address in the future. We also decided the simple linear model was the best fit for the maximum temperature data. Using this model, we found significant evidence for the increase in temperature over time (for every 1 additional year, the temperature increased 0.0237 degrees Celsius, on average), and the model was able to explain 33.9% of the variation in temperatures through the relationship to time.

We would like to get more detailed information from the Togolese meteorologists before we proceed with our research. To gain a better understanding of the data we are working with, we would like to have more details on how and when (how often?) the data was collected, as well as who was responsible for the collection. We would like to get more information about the two years (1968 and 1972) that we found to be outliers in the maximum temperature data. We are interested in whether any abnormal events (i.e. a drought) around Atakpamé could have possibly occurred during these years, as this would explain the discrepancy between these points and the rest of the annual averages.