**Name: Deniz Sert**

**Kerberos: dsert**

Problem Set 5 Write-up

Modeling Temperature Change

# Problem 4

Plot 4A: Average Daily Temperature for Portland on 12/25 (1961-2016)

|  |
| --- |
|  |

Plot 4B: Average Yearly Temperature for Portland (1961-2016)

|  |
| --- |
|  |

4.1 What difference does choosing a specific day to plot the data versus calculating the yearly average have on the goodness of fit of the model? Interpret the results.

|  |
| --- |
| Plotting the yearly average instead of the specific days increased the goodness of fit of the model. The Standard error decreased from 11.66 to 0.198, indicating that the yearly average was a better fit for the model. |

4.2 Why do you think these graphs are so noisy?

|  |
| --- |
| The graphs are noisy due to random variation from the true underlying pattern. This is typical in statistical models. |

# 

# Problem 5

Plot 5B: Increasing Interval (Phoenix, length=30)

|  |
| --- |
|  |

5.1 What were the start and end years for your window[[1]](#footnote-1)? What was the slope?

|  |
| --- |
| Start : 1962, End : 1992  Slope : ﻿0.1404 |

Plot 5C: Decreasing Interval (Phoenix, length=15)

|  |
| --- |
|  |

5.2 What were the start and end years for your window[[2]](#footnote-2)? What was the slope?

|  |
| --- |
| ﻿Start : 1985, End : 2000  Slope: -0.0797 |

5.3 Considering *both* plot 5B and 5C, what conclusions, if any, can you make with respect to how temperature is changing over time based on your models and their goodness of fit?

|  |
| --- |
| Since the difference between Start and End for the positive slope was 15 degrees greater than the difference between Start and End for the negative slope, there is a larger interval for the increasing temperature.  In addition, the Standard Error for the increasing interval, 0.1018, is greater than the Standard Error for the decreasing interval, -0.2826. This means that the increasing interval model has a better fit than the decreasing interval model.  Therefore, the increasing temperature model fits the data better than the decreasing temperature model. |

# Problem 6

Plot 6B: Training Data (1961-1999), Degree 2

|  |
| --- |
|  |

Plot 6B: Training Data (1961-1999), Degree 10

|  |
| --- |
|  |

6.1 How do these models compare to each other in terms of R^2 and fitting the data?

|  |
| --- |
| The degree 10 model has a greater R^2 value than the degree 2 model.  In degree 2, 76.5% of the variation in temperature can be explained by time, whereas in degree 10, 77.5% of the variation in temperature can be explained by time.  Therefore, the degree 10 model fits the data better than the degree 2. |

Plot 6B: Test Data (2000-2016), Degree 2

|  |
| --- |
|  |

Plot 6B: Test Data (2000-2016), Degree 10

|  |
| --- |
|  |

6.4 Which model performed the best and how do you know? If this is different from the training performance in the previous section, explain why this occurred.

|  |
| --- |
| The degree 2 model, with an RSME of 0.26, performed better than the degree 10 model, with an RSME of 0.24.  This is a different result than the training data because of the natural noise that occurs with random data. The degree 10 model fit the noise better, resulting in performance in the training data, but was not able to perform well for the testing data.  Although the degree 2 model did not fit the training data as well as the degree 10 model, it fit the testing data better because it fit the underlying model rather than the noise. |

6.5 If we had generated the models using the data from Problem 4B (i.e. the average annual temperature of Portland) instead of the national annual average over the 22 cities, how would the prediction results on the national data have changed?

|  |
| --- |
| The prediction results would not perform as well.  The data from one city is not representative of the national average.  In the Portland example, more rain would cause the temperature to decrease, while in Reno, the intense sun would cause the temperaure to increase. This is an example of a lurking variable. |

1. The end year should be the last year whose temperatures are included in the model. [↑](#footnote-ref-1)
2. The end year should be the last year whose temperatures are included in the model. [↑](#footnote-ref-2)