**Least absolute deviations (LAD) -- A simple look into Climate Data Analysis**

**Name: Zonghao Zou (**[**zzou8@wisc.edu**](mailto:zzou8@wisc.edu)**) 9075215013**

**Zirui Jiang (**[**zjiang75@wisc.edu**](mailto:zjiang75@wisc.edu)**) 9073538386**

**Yuhan Xu (**[**yxu329@wisc.edu**](mailto:yxu329@wisc.edu)**) 9075265950**

**Group number: 32**

**Proposal**

We will introduce Least Absolute Deviation (LAD) that uses the mathematical foundation of ordinary least squares we covered in class. Least absolute deviations is a statistical optimization technique that relies on the L\_1 norm condition. It attempts to find a function that closely approximates a set of data. The main benefit of using LAD over least squares is that the training result is insensitive to “outliers”.

The purpose of this project is to present LAD to students and offer them an alternative method of doing data analysis. For which, the expression can be written as the following:

**Learning Objectives**: In this project, we will focus on a paper published by Vanderbei named “Local Warming” We will build out activities around the model presented in this paper and provide students an opportunity to interact with a real life model. Throughout the process, we hope that students will gain a better understanding of:

1. how LAD functions, where the climate model calculated several key features

2. the benefits of LAD, insensitive to outliers, and generates more believable results

3. the major difference between LAD and least squares

4. doing simple climate data analysis and solving linear programming problem

A least-squares model failed to characterize the solar cycle correctly.

bring a new alternative statistical approach to these analyses that should allow one to extract meaningful climate information from noisy data in a more robust and reliable manner than before.

minimize the sum of the absolute deviations

We will use 50 years data of daily average temperature from a local weather station. Since a least-squares model fails to characterize the solar cycle correctly and is sensitive to outliers in the data, we make a least-absolute-deviations regression model to find the values of unknown regression coefficients that minimize the sum of the absolute deviations. In particular, we are going to be graphly see the trend of nominal temperature, compute the amplitude of annual seasonal changes in temperatures, and compute the amplitude of the temperature changes brought about by the solar-cycle. The least absolute deviation regression was introduced in 1757 by Roger Joseph Boscovich. He used this procedure while trying to reconcile incoherent measures that were used to estimate the shape of the earth. Least absolute deviations is very important because it is robust in that it is resistant to outliers in the data. LAD gives equal emphasis to all observations, in contrast to ordinary least squares (OLS) which, by squaring the residuals, gives more weight to large residuals, that is, outliers in which predicted values are far from actual observations. LAD is interesting to study because LAD overcomes the aforementioned drawbacks of the least squares regression and provides an attractive alternative. It is less sensitive than least squares regression to the extreme errors and assumes absolute error loss function. Because of its resistance to outliers, it provides a better starting point than the least squares regression for certain robust regression procedures. Unlike, other robust regression procedures, it does not require (a rejection parameter).

This may be helpful in studies where outliers do not need to be given greater weight than other observations.

<https://stats.stackexchange.com/questions/313235/can-i-use-gradient-descent-for-least-absolute-deviation-regression>

<https://stats.stackexchange.com/questions/240804/r2-for-least-absolute-deviation-regression>