The Koester Equation: The Quantification of Progress Lost in a Data Set

By **Stefan Wolfgang Koester**

ORCID #: 0000-0003-1156-4090

Abstract

The Koester Equation is the mathematical processes that allows for the quantification of a negative trend that is able to express numerous precious qualities about the data given. Uniquely, the Koester Equation conforms to fit all possibilities of data given. The Koester Equation can even be taken to another level and be used as a relative measure, showing how certain pattern and trends appear because of different human variables. The Koester Equation, and all of its processes, quantify the "loss in progress" experienced in a data set when it undergoes an abnormality, such as a missing day in testing. This loss in progress can also be viewed as a number determining by how much that data set is skewed by an abnormality. For example, if a person were to take three of the same tests for three days in a row, an obvious positive curve in their results would be apparent. If, on the fourth day, a break was taken and no testing occurred, the results after would not be the same as if the person had just continued. This is usually known as the loss in progress, and can now be quantified using The Koester Equation.

Keywords: statistics, Koester, equation, data, analysis, analytics, linear, quadratic, regression, method, system, progress, inconsistency, curve

1. Introduction

According to Samprit Chatterjee's book on regression analysis, regression analysis has numerous areas of applications.[1] Imagine seven or more values in slight progression. Originally, if one of the values were not to exist, the natural progression of data would be skewed. **Using The Koester Equation, one is able to determine by how much a data set is skewed by an inconsistency in data (in terms of progress).** The equation uses values obtained from a graph with quadratic and/or line regressions formed from separate value sets. The value obtained can also be seen as the amount of progress lost (loss of progress) because of the inconsistency. Through this process, the quantification of progress lost in a data set because of an abnormality or absence in a data set is possible.

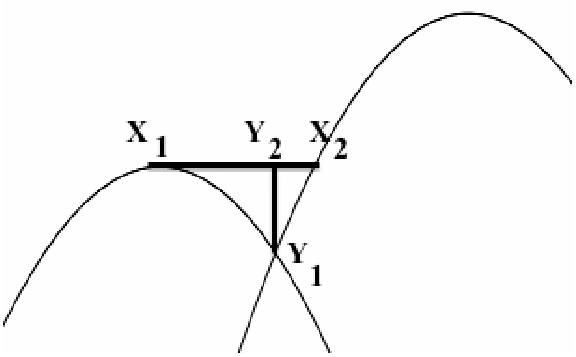
2. Methods

Both sets of values (minimally three each) create two regressions (that is able to be obtained by calculation or a calculator). These regressions (either quadratic or linear) are chosen by the curve of the data points. If there is an obvious curve, a quadratic regression best fits the data. If no curve exists, then a line regression is most appropriate. For the most accurate decision on the type of regression, use the method described below under "Curve-Fitting." If a straight line is drawn from the vertex (highest point) of the first quadratic regression to the second regression, the length of the line would obviously be the the change of X coordinates from these two points. This line creates a shape similar to a "v." If another line is then drawn

[1] Chatterjee, Samprit, and Bertram Price. "Publicly Available Data Sets." *Regression Analysis by Example*. New York: Wiley, 1977. N. pag. Print.

perpendicularly from the lowest point in this shape (the bottom of the "v") to the straight line above, this value is the height. With these two values, the loss of progress can be calculated using the equation. The basic template is shown below.

The general template above shows how the values for the equation are attained. The basic equations (with regards to the template) are:



With this, the equation is:

$$P_l = \left| \frac{h}{\Delta x} \right| = \left| \frac{(Y_2 - Y_1)}{(X_2 - X_1)} \right|$$

2A. Curve-Fitting

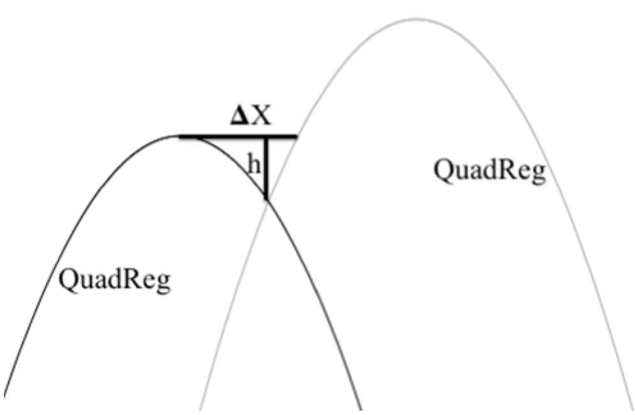
How can one know whether to use a quadratic regression or a linear regression? With the three data points, the slopes are the indicator for the type of regression. In the slopes in the data points, the denominator will always be one (for the change in X values will always be one), while the numerator will be the difference in the Y values. So, to chose whether to use a linear regression or a quadratic regression, one must compute the two slopes between the points. If the slopes are similar, then a linear regression best fits. If the slopes are different, then a quadratic regression best fits. To determine by how much they are different, take the difference in the two slopes and divide it by the slope of the first two points. It will always be a number less than one. If the number is greater than 0.20, use a quadratic regression. If the number is less than 0.20, use a linear regression, for the data points are too similar. The type of regressions and how they are used becomes a huge part in the process as it pertains to creating the statistical model needed to solve for the purpose. As Dr. Freedmen expresses in his book, statistical models place emphasis on the connections occurring between the models and real phenomena.[2]

Rules of the Equation:

• The two calculated regressions must intersect

• One regression needs to be a quadratic regression • All trends in scores must be positive

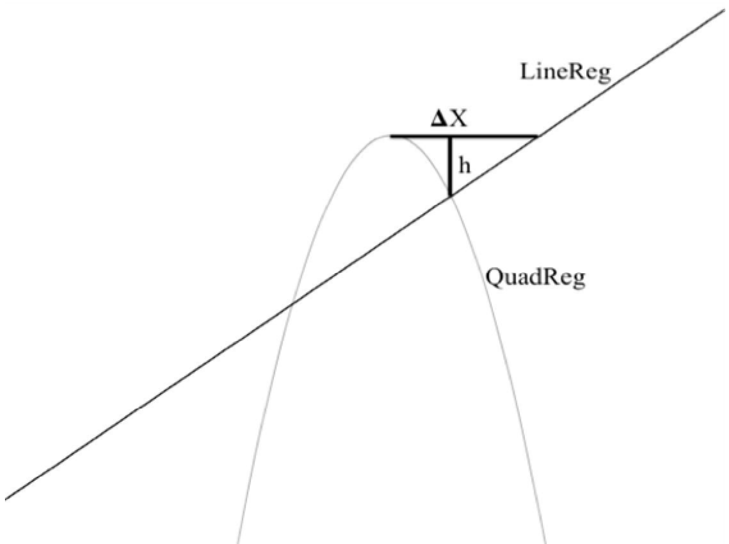
Type 1: 2 Quadratic Regressions



Type 2: One Linear Regression, One Quadratic Regression

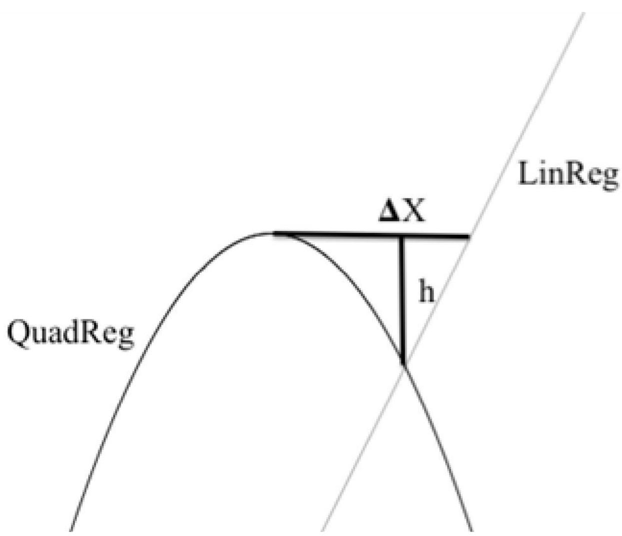
[2] Freedman, David. "Preface." Preface. *Statistical Models: Theory and Practice*. Cambridge: Cambridge UP, 2005. N. pag. Print.

• Only works when the fourth value (fifth day) in the QuadReg falls below the regular line regression



Type 3: QuadReg, LinReg

• the first three values represent a QuadReg while the remaining values represent a LinReg



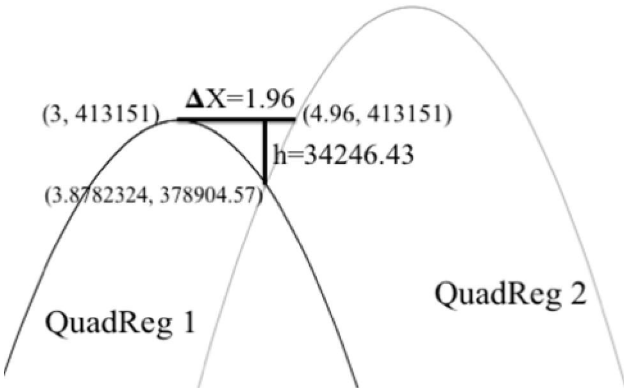
3. Evaluation

All of the methods and equations have been tested thoroughly in many real situations. The only times where the Koester Equation failed to compute a meaningful number was when the day of rest actually caused a spike in scores on the fifth day and the following days. On this occasion, it is understood that the number would become zero, because no progress was lost because of the absence of one day, and the data was not skewed negatively. Other times, the absence of a day causes negative spike, where numbers fall far below its precedents. In this case, the equation and methods still apply, and will compute a meaningful number. Many question the use of regressions to compute such a precise number. Those people fail to take into account the accuracy of the regression model as a whole. Also, since the research and mathematics behind The Koester Equation pertain to test results and other data sets acquired from human test results, it can be considered social research. In accordance with John Fox's beliefs stated in his book, linear models and regressions are the most useful and widely used tools for social research.[3]

4. Illustration

On day one, an individual scores 271,681 on a certain anatomy. For the next two days, she scores 380,540 and 413,151. These three values best represent a quadratic regression, or QuadReg. After the third day of testing, the individual stops, and does not take the test until the fifth day, where she scores a 414,544. For the next couple of days, she scores a 440,000; a 463,271; and a 479,200. This also represents a quadratic regression. Using the Koester Equation, one can measure the loss of progress Grace experienced by the absence in test taking on the fourth day can be calculated. Refer to the graph below for a visual.

[3] Fox, John. *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: Sage, 2008. Print.



Work Shown:

$$P_l = \left| \frac{h}{\Delta x} \right| = \left| \frac{(Y_2 - Y_1)}{(X_2 - X_1)} \right|$$

$$P_l = \left| \frac{34246.43}{1.96} \right|$$

$$P_l = 17472.66837$$

4A. Explanation

The first three values create a quadratic regression, and the next three also show a quadratic regression. They intersect so that the two needed values are easily acquired. Drawing a line from the vertex of the first quadratic regression right to the side of the second quadratic regression with the same y value, yields a value of 1.96 for the "change in x value." Now, creating a line from the primary line connecting the two regressions to the intersection point (the line must be perpendicular to the primary line), a value is created from the length of that line, being 34246.43. According to the equation, the loss of progress is equal to the absolute value of the quotient of dividing the length of height (or length of secondary line) by the "change of x" (length of primary line). This is 34246.43 divided by 1.96, which is 17472.66837. So, the loss of progress experienced by the absence of one test is the deficit of 17472.66837.

5. Discussion

The Koester Equation and the methods behind it are amongst many mathematical ways to obtain a measure of how much a data set is skewed. However, the Koester Equation is the first to use horizontal and vertical measures between two regressions to analyze the progress lost due to a natural absence in data, and do so in a way that fits to most situations. This is very helpful in analyzing data, and viewing trends in improvement and retrogression.

The Koester Equation and all of its processes are among the select mathematical processes that are able to quantify a negative trend that is able express numerous precious qualities about the data given. Uniquely, the Koester Equation conforms to fit all possibilities of data given, which is especially admirable because the human brain is capable of any type of results, and to have a mathematical process quantify a measure of loss in any type of results is unique. The Koester Equation can even be taken to another level and be used as a relative measure, showing how certain pattern and trends appear because of different human variables.

Acknowledgments

[1] Chatterjee, Samprit, and Bertram Price. "Publicly Available Data Sets." *Regression Analysis by Example*. New York: Wiley, 1977. N. pag. Print.

[2] Freedman, David. "Preface." Preface. *Statistical Models: Theory and Practice*. Cambridge: Cambridge UP, 2005. N. pag. Print.

[3] Fox, John. *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: Sage, 2008. Print.