2009

Repeatability Estimates Of Sloped Scattered Data

Angelique Waller
University of South Florida

Advisors:
Arcadii Grinshpan, Mathematics and Statistics
Alex Volinsky, Mechanical Engineering

Problem Suggested By: Alex Volinsky

Follow this and additional works at: https://digitalcommons.usf.edu/ujmm

Part of the Mathematics Commons

UJMM is an open access journal, free to authors and readers, and relies on your support:
Donate Now

Recommended Citation
Waller, Angelique (2009) "Repeatability Estimates Of Sloped Scattered Data," Undergraduate Journal of Mathematical Modeling: One + Two: Vol. 2: Iss. 1, Article 8.
DOI: http://dx.doi.org/10.5038/2326-3652.2.1.8
Available at: https://digitalcommons.usf.edu/ujmm/vol2/iss1/8
Repeatability Estimates Of Sloped Scattered Data

Abstract
Repeatability is the variance in data accumulated under fixed conditions. It is important for quality control as it costs both time and money to recalibrate tools and remanufacture machines. This project compares three methods for approximating the repeatability of a sloped scattered data set. The first method uses a linear approximation, the second involves rotating the data points, and the third calculates distance using right triangles. The methods are compared for both precision and ease of use.

Keywords
Repeatability, Standard Deviation, Quality Control, Sloped Data

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.
TABLE OF CONTENTS

Problem Statement................................................................. 3

Motivation................................................................................. 3

Mathematical Description and Solution Approach...................... 4

Discussion............................................................................... 9

Conclusion and Recommendations............................................ 11

References............................................................................... 12

Appendices.............................................................................. 13-15
PROBLEM STATEMENT

Repeatability is the measure of change in the variance of an operation performed repetitively under controlled conditions. Calculation of repeatability can determine when a machine is malfunctioning.

Suppose a robotic arm is given the instructions to weld along a sloped line. After the machine runs, coordinates for the position of the weld are measured at regular intervals. The coordinates for two welds are $W_1 = (X_1, Y_1)$ and $W_2 = (X_2, Y_2)$. Derive a formula for repeatability and explore various ways to apply this formula for sloped data. Compare the various techniques for both ease and accuracy of computation.

MOTIVATION

Any machine used in production will begin to wear down from the constant repetition of tasks. After enough time the internal parts of any machine begin to fatigue causing the machined parts to repeatedly fail inspection by quality control. Once a machine has reached this point, no amount of recalibration will allow it to run correctly. However, not all machine error is caused by faulty parts.

Sometimes when a machine is bumped or moved, it will lose its point of reference. Though the machine is mechanically functional, the work will still contain errors. In this case, a simple recalibration will allow the machine to find its point of reference and it will produce accurate work again. Precision in both machines and instruments is critical to manufacturing. Statistics such as repeatability are useful in determining whether a machine needs to be remanufactured or simply recalibrated.
MATHEMATICAL DESCRIPTION AND SOLUTION APPROACH

REPEATABILITY AND STANDARD DEVIATION

Repeatability measures the change in variance of a repeated task under fixed conditions. In statistics the standard deviation is the square root of the variance, so a study of variance is a study of the standard deviation. For a set \( D = \{d_1, d_2, \ldots, d_n\} \), an unbiased estimator for the standard deviation \( \sigma_y \) is

\[
s(D) = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}
\]

where \( \bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i \). We wish to express repeatability as a percentage in terms of the standard deviation of the two set being compared, namely \( D_1 \) and \( D_2 \). The percentage change in standard deviation is thus

\[
\frac{|s(D_1) - s(D_2)|}{\max\{s(D_1), s(D_2)\}}.
\]

Rationally, 100% repeatability should represent that \( s_1 = s_2 \), and 0% repeatability should mean that either \( s_1 \gg s_2 \) or \( s_1 \ll s_2 \). With this in mind, repeatability will be defined as

\[
Rep(D_1, D_2) = 1 - \frac{|s(D_1) - s(D_2)|}{\max\{s(D_1), s(D_2)\}} = \begin{cases} s(D_2) / \sigma(D_1), & s_1 \geq s_2 \smallskip \sigma(D_1) / s(D_2), & s_1 < s_2 \end{cases}.
\]
**Repeatability of Sloped Data**

In the case of sloped data, a pair of coordinates \((x_i, y_i)\) is given rather than just a single value \(d_i\). As a result, the estimator for the standard deviation in (1) cannot be computed. This is the fundamental problem explored throughout the remainder of this project.

To remedy the situation, each coordinate \((x_i, y_i)\) is evaluated as a distance \(d_i\) to the intended trend-line. If the theoretical trend-line is known explicitly it should be used throughout these calculations. If the theoretical line is not known explicitly, the trend-line can be approximated through linear regression of the first data set as

\[
y = \overline{\beta}_1 x + \overline{\beta}_2
\]

where

\[
\overline{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \quad \text{and} \quad \overline{\beta}_2 = \overline{y} - \overline{\beta}_1 \overline{x}
\]

such that \(\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i\) and \(\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i\). For the remainder of the paper, the exact or approximated trend-line will be given by the function

\[
f(x) = ax + b.
\]

We now explore several ways to define a distance function, \(dist\), from a point \((x, y)\) to the line \(f\).
Method 1: Linear Approximation

The first method explored for calculating distance is a linear (vertical) approximation,

\[ \text{dist}(x_i, y_i, f) = y_i - f(x_i) \]  \hspace{1cm} (3)

This method simply subtracts the measured value of \( y_i \) from the expect value of \( y_i \) at \( x_i \). With this method, computation of the distance is quick and straightforward.

Method 2: Rotation

In the rotation method, the \( xy \)-plane is rotated along around the origin until the slope of the trend-line is zero. For ease of computation, after the rotation is performed we will shift the trend-line to the axis by subtracting everything by the \( y \)-intercept value. The linear transformation which rotates a point \((x_i, y_i)\) by \( \alpha \) radians in the clockwise direction is

\[
\begin{pmatrix}
  x_i' \\
  y_i'
\end{pmatrix} = \begin{pmatrix}
  \cos(-\alpha) & -\sin(-\alpha) \\
  \sin(-\alpha) & \cos(-\alpha)
\end{pmatrix} \begin{pmatrix}
  x_i \\
  y_i
\end{pmatrix}.
\]  \hspace{1cm} (4)
Using matrix multiplication and the fact that \( \cos(-\alpha) = \cos(\alpha) \) and \( \sin(-\alpha) = -\sin\alpha \), line (4) becomes the system

\[
\begin{align*}
&x'_i = x_i \cos \alpha + y_i \sin \alpha \\
y'_i = -x_i \sin \alpha + y_i \cos \alpha \tag{5} \\
&\tag{6}
\end{align*}
\]

If we let \( \alpha = \tan^{-1} a \), the rotation will cause the trend-line to become horizontal, i.e. the slope of the new rotated line \( \tilde{f} \) will equate to zero. To get the precise equation for the rotated trend-line, simply track any point on \( f \) through the rotation transformation. For instance, consider the \( y \)-intercept, which for \( f \) is \((0, b)\). Thus from (5) and (6), \( x' = b \sin \alpha \) and \( y' = b \cos \alpha \). Since we know that \( \tilde{f} \) is horizontal and passes through \((x', y')\) its formula must be

\[
\tilde{f}(x) = b \cos \alpha. \tag{7}
\]

To simplify the distance calculation, the rotated data points and trend-line can be shifted to the \( x \)-axis by subtracting \( b \cos \alpha \). At this point the distance from the transformed points to the transformed trend-line is simply the \( y \)-value of the transformed point, i.e.

\[
dist((x_i, y_i), f) = -x_i \sin \alpha + (y_i - b) \cos \alpha. \tag{8}
\]
Method 3: Minimum Distance

This method uses trigonometry to compute the shortest distance between a point and the trend-line. Consider the following picture (Picture 3),

For any point \((x_i, y_i)\), there is a line running through this point which is perpendicular to the trend-line. The trend-line, the perpendicular line, and the vertical line through the point form a right triangle. The Pythagorean Theorem says that the hypotenuse (the vertical segment) will always be longer than either of the legs (the trend-line and corresponding perpendicular segment). This means that the shortest distance from the point to the trend-line is the segment perpendicular to the trend-line.

If the line increases at an angle of \(\alpha\), the length of the perpendicular segment is given by the following relationship between \(\alpha\) and the hypotenuse (the vertical segment):

\[
\frac{\text{Perpendicular}}{y_i - f(x_i)} = \frac{\text{Perpendicular}}{\text{Hypotenuse}} = \cos \alpha
\]  

(9)
which simplifies to

\[ \text{dist}(x_i, y_i, f) = (y_i - f(x_i)) \cos \alpha. \]  

(10)

**DISCUSSION**

Each of the three methods were used to evaluate two sets of welding data $W_1 = (X, Y_1)$ and $W_1 = (X, Y_2)$; see Table 1 in Appendix A. Excel was used to plot the data and compute the trend-line (Chart 1, Appendix A) to calculate the repeatability. From the table, it is clear that all three methods calculate a 95% repeatability for the welding machine. Since all three methods yield the same result, the most efficient statistic to use is the easiest to compute. Out of the three methods, the linear approximation method is easiest to compute and should be favored.
Intuitively it would seem that the linear method would give an inaccurate measurement of repeatability since the distances from the trend-line to the data points are not optimized, but upon further inspection:

\[
\begin{align*}
    s_{\text{Min}}(W) &= \sqrt{\frac{\sum_{i=1}^{n}(\cos \alpha(y_i - f(x_i)) - \sum_{i=1}^{n} \cos \alpha(y_i - f(x_i)))^2}{n-1}} \\
    &= \cos \alpha \sqrt{\frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{n-1}} \\
    &= \cos \alpha \cdot s_{\text{Lin}}(W)
\end{align*}
\]

so the repeatability using the minimum distance is

\[
\frac{Rep_{\text{Min}}(W_1, W_2)}{s_{\text{Min}}(W_2)} = \frac{s_{\text{Min}}(W_1)}{s_{\text{Min}}(W_2)} = \cos \alpha \cdot s_{\text{Lin}}(W_1) = Rep_{\text{Lin}}(W_1, W_2)
\]

which is the same as the repeatability using the linear approximation for \( s_{\text{Min}}(W_1) < s_{\text{Min}}(W_2) \).

Similarly, \( Rep_{\text{Min}}(W_1, W_2) = Rep_{\text{Lin}}(W_1, W_2) \) for \( s_{\text{Min}}(W_1) \geq s_{\text{Min}}(W_2) \).

Geometrically, the minimum distance and rotational method can be seen to give the same distance. By rotating the trend-line so that it is horizontal, the segment perpendicular to the trend-line will be vertical. Since the rotation has not changed the distance between the point and the trend-line, the length of the perpendicular segment will be the same throughout the rotation. Thus the rotation method and minimum distance method will give the same distance. If both methods give the same distance values, they will also give the same standard deviations and, by implication, the same repeatability.
CONCLUSION AND RECOMMENDATIONS

The repeatability is defined as the percentage change in variance of the measurements under controlled conditions. Identifying and minimizing error in experiments or manufacturing is crucial because it costs time, money and effort to have a machine remanufactured when it can simply be recalibrated. This study was important because there will always be numerous methods for accomplishing the same task but efficiency will render some methods obsolete.

In this problem, three methods to compute the repeatability for sloped data were explored. The first method was a linear approximation where the vertical distance from the point to the trend-line was the only consideration. The second method was a rotation and shift so that the trend-line would lie on the \( x \)-axis. The third method involved using trigonometry to find the minimum distance between a point and a line. In the end, each method was shown to give the same computation of repeatability, but the linear approximation should be favored for its ease of computation.

For the welding data, the machine was shown to have a 95% repeatability which is acceptable. Continued measurements of the reparable will indicate when the machine needs to be maintained.

Future projects could explore the repeatability calculations for objects in 3D. Many machines manufacture goods using measurements in 3D and the repeatability calculations will surely be more difficult than the cases considered here.
REFERENCES

Frank, Harry, and Steven C. Althoen. *Statistics concepts and applications*. New York: Press Syndicate of the University of Cambridge, 1994. Print.

Neuilly, Michele. *Modelling and Estimation of Measurement Errors*. Null: Lavoisier, 1999. Print.

Stevens, James P. *Intermediate Statistics A Modern Approach*. 3rd ed. Mahwah: Lawrence Erlbaum, 2007. Print.
### APPENDIX A – TABLE AND CHART

| Measured Values | y-(Trend) | Distances |
|-----------------|-----------|-----------|
|                | Linear    | Rotation  | Minimum |
|                | L₁, L₂    | R₁, R₂   | D₁, D₂  |
| X               |           |           |         |
| 0.00            | 0.06140   | 0.05749  |          |
| 0.05            | 0.07264   | 0.07115  |          |
| 0.10            | 0.06608   | 0.05560  |          |
| 0.15            | 0.05860   | 0.05706  |          |
| 0.20            | 0.06235   | 0.08723  |          |
| 0.25            | 0.08117   | 0.07125  |          |
| 0.30            | 0.09235   | 0.07927  |          |
| 0.35            | 0.09142   | 0.10327  |          |
| 0.40            | 0.09151   | 0.08579  |          |
| 0.45            | 0.10190   | 0.10840  |          |
| 0.50            | 0.11594   | 0.10134  |          |
| 0.55            | 0.12433   | 0.12140  |          |
| 0.60            | 0.12908   | 0.12867  |          |
| 0.65            | 0.13640   | 0.11755  |          |
| 0.70            | 0.13446   | 0.13927  |          |
| 0.75            | 0.12984   | 0.15211  |          |
| 0.80            | 0.12818   | 0.15273  |          |
| 0.85            | 0.13769   | 0.15980  |          |
| 0.90            | 0.15180   | 0.16376  |          |
| 0.95            | 0.16480   | 0.14727  |          |
| 1.00            | 0.17236   | 0.17415  |          |
| 1.05            | 0.18059   | 0.18358  |          |
| 1.10            | 0.18236   | 0.16564  |          |
| 1.15            | 0.18129   | 0.18807  |          |
| 1.20            | 0.17683   | 0.18508  |          |
| 1.25            | 0.17894   | 0.20647  |          |
| 1.30            | 0.18764   | 0.20548  |          |
| 1.35            | 0.20303   | 0.20123  |          |
| 1.40            | 0.22510   | 0.22575  |          |
| 1.45            | 0.23478   | 0.20814  |          |
| 1.50            | 0.23192   | 0.22253  |          |

| Calculation of the repeatability based on two measured data sets using the linear, rotation, and minimum distances. |
### Table 1 (cont.): Calculation of the repeatability based on two measured data sets using the linear, rotation, and minimum distances.

| Measured Values | y-(Trend) | Linear | Rotation | Minimum |
|-----------------|-----------|--------|----------|---------|
| X               | Y₁        | Y₂     | L₁       | L₂      | R₁  | R₂  | D₁  | D₂  |
| 1.55            | 0.22863   | 0.22073| 0.22823  | 0.05394 | 0.04604 | 0.00040 | -0.00745 | 0.00040 | -0.00745 |
| 1.60            | 0.23446   | 0.23382| 0.23387  | 0.05413 | 0.05350 | 0.00059 | -0.00004 | 0.00059 | -0.00004 |
| 1.65            | 0.24316   | 0.22501| 0.23950  | 0.05720 | 0.03905 | 0.00363 | -0.01440 | 0.00363 | -0.01440 |
| 1.70            | 0.23946   | 0.25279| 0.24514  | 0.04786 | 0.06119 | -0.00564 | 0.00760 | -0.00564 | 0.00760 |
| 1.75            | 0.23738   | 0.24226| 0.25077  | 0.04015 | 0.04503 | -0.01331 | -0.00846 | -0.01331 | -0.00846 |
| 1.80            | 0.24904   | 0.24970| 0.25641  | 0.04617 | 0.04684 | -0.00732 | -0.00666 | -0.00732 | -0.00666 |
| 1.85            | 0.26661   | 0.27361| 0.26204  | 0.05811 | 0.06511 | 0.00454 | 0.01150 | 0.00454 | 0.01150 |
| 1.90            | 0.27825   | 0.26054| 0.26768  | 0.06411 | 0.04640 | 0.01051 | -0.00710 | 0.01051 | -0.00710 |

| Standard Deviation: | 0.00902  | 0.00949  | 0.00897  | 0.00943  | 0.00897  | 0.00943  |
| Repeatability:       | **0.94787** | **0.94787** | **0.94787** |

**Chart 1:** Data collected from two welds run on the same machine under the same conditions and the line which best fits the first weld.

\[ y = 0.112x + 0.053 \]
APPENDIX C - FORMULAS

DISTANCE CALCULATIONS

Consider the following distance formulas from a point $(x_i, y_i)$ to a line $f(x) = ax + b$.

Let $\alpha = \tan^{-1} \alpha$.

**LINEAR APPROXIMATION:**

$$\text{dist}((x_i, y_i), f) = y_i - f(x_i)$$

**ROTATION:**

$$\text{dist}((x_i, y_i), f) = -x_i \sin \alpha + (y_i - b) \cos \alpha$$

**MINIMUM DISTANCE:**

$$\text{dist}((x_i, y_i), f) = (y_i - f(x_i)) \cos \alpha$$