The Oval Experiment, Measuring Distance

Rudy A. Gideon
University of Montana
Missoula, MT 59812

Abstract

An actual experiment in an applied statistics class is beneficial in helping students understand the many complexities of estimation. In a senior- graduate level applied statistics class the following independent two-sample experiment was carried out by the students. An Oval sits in the heart of the University of Montana campus with only grass and sidewalks on the edge and interior of this Oval. Bordering on this Oval is the mathematics building in which the classes are held. The distance around this Oval is taken as an unknown parameter and many people walk it every day. There are two classes of students and each class was assigned a different way to estimate the distance. The questions of interest are; (1) are the two methods equally valid (an independent two-sample question), and (2) what statistic best estimates the distance around; a robust estimation procedure or a classical one. The main emphasis is the introduction to robust estimation through rank based correlation coefficients, in particular the Greatest Deviation correlation. A general method of estimation with any correlation coefficient is demonstrated with a rank based correlation. Location, scale, and simple linear regression parameters are estimated and compared to the classical estimators. The simplicity and usefulness of the new method is apparent.

Keywords: two-sample problem, scale estimation, robust estimation, Greatest Deviation Correlation, robust location estimation

AMS Subject Classification: 62

1. The experimental methods

The first class had to measure the distance by a timer method. Just north of the campus next the Clark Fork River is a 400-meter track. The students were to time themselves walking around the 400-meter track and then time themselves walking the Oval (hopefully at the same pace). They then could estimate the Oval distance by comparing the time ratios to the known and unknown distances.

The second class used the same 400-meter standard, but instead counted steps. These students then could estimate the distance by comparing the step ratio to the known and unknown distances.

Each student in each class performed their own experiment and thus, calculated their own estimate. The timer method had 24 independent observations and the count method 25. Before comparing classes, standard statistical techniques were used to evaluate the quality of the data. Boxplots revealed that 5 to 8 of the observations were “outliers.”

The count method probably had 4 outliers while the timer method had 3. The sample standard deviations were around 50 meters with the outliers in and each dropped to about 22 meters when the few outliers were deleted.

The data is listed at the end and there are three graphs with a robust analysis that used the Greatest Deviation Correlation Coefficient (GD). The first graph plots a normal Quantile plot of both data sets with a GD regression fit. The next two, GD-fit, quantile plots are for the pooled data; the first with all the data and the second with possible outliers deleted.

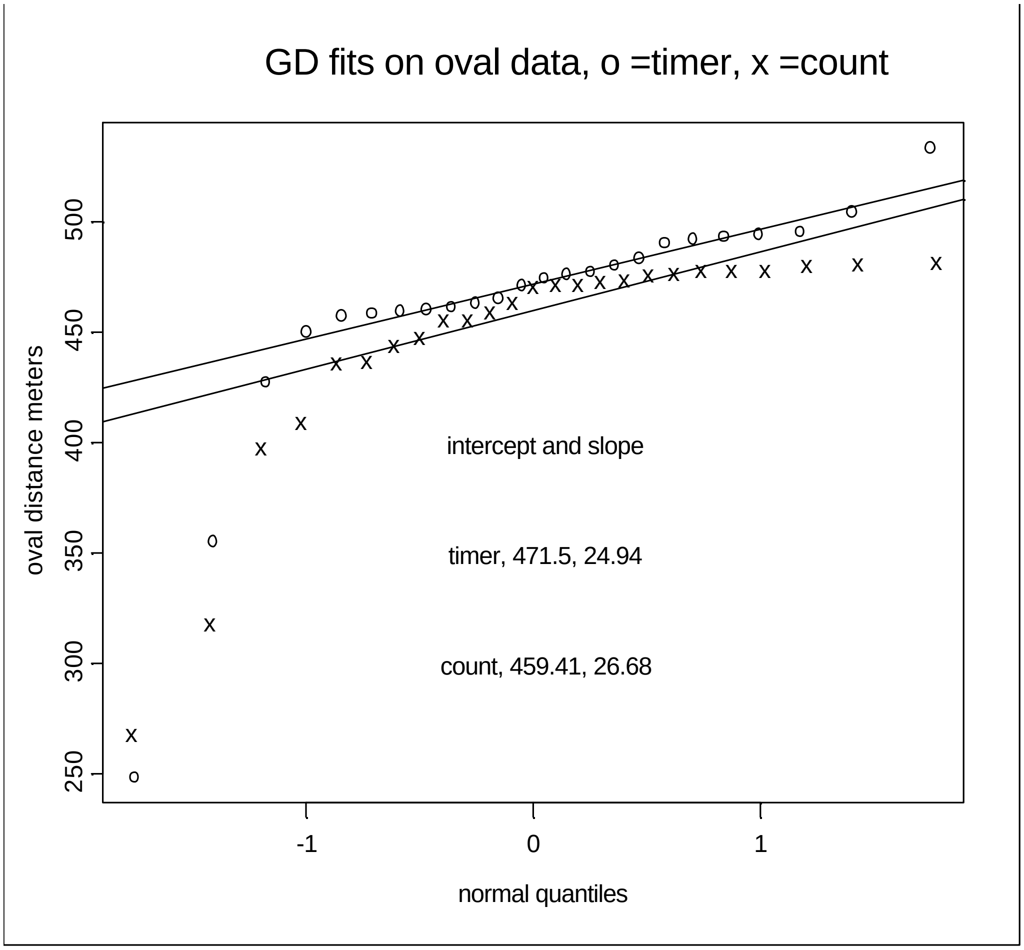
The quickest way to evaluate the data is to use the GD quantile plot in which the ordered combined data is plotted against standard normal quantiles. The Greatest Deviation correlation method is used to fit a regression line to the bulk of the data. It is clearly seen that most of the data follow the normal but there are outliers at each end. There is only one possible outlier on the upper end and the others are in the low range. The estimate of the mean is 466.33,the intercept, and the standard deviation estimate is 21.52, the slope. These estimates come without any energy spent deciding what is good and what is bad data. With trimmed means or deleting outliers via the interquartile method, one can get these GD values. However, what percentage should be used for the trimmed means? How many data points should be deleted? Depending on one's choice different estimates are obtained. With GD there are no choices, one gets reasonable estimates on all of the data. If one uses the interquartile 1.5 rule for outliers on the pooled data there are 6 outliers of low values and one upper outlier. An outlier is a point more than 1.5 times the IQR (interquartile range) above or below the first and third quartiles. These points are seen on the pooled normal quantile plot as large deviations from the GD line fit.

The data are given in the table below. The first column contains the timer data in the first 24 rows and the last 25 rows contain the count data. The second column contains the pooled ordered data with the 3rd column providing the data type with 0 being the timer and 1 for the count method. The 4th column provides a count.

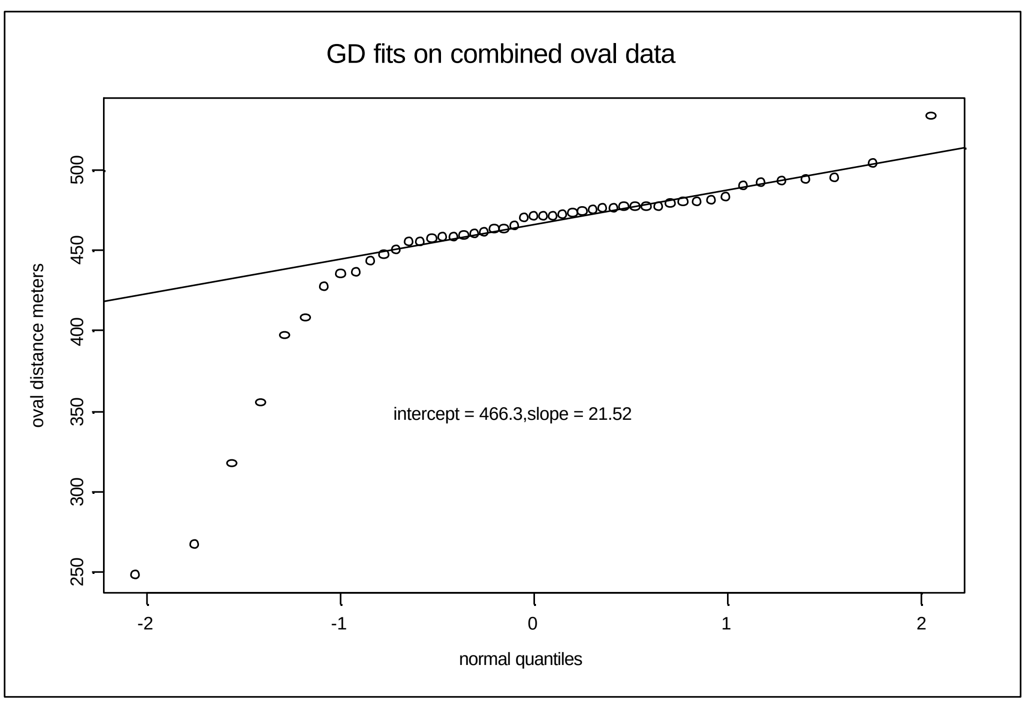
In the two graphs that are below, the robustness of GD is shown in location and scale estimates. First separate plots of the timer and count data appear on the first quantile plots. The robust GD means are 471.50 and 459.41 which compare to the actual means of 461.04 and 446.40. The GD estimates of standard deviation are 24.94 and 26.68 which compare the classical sample standard deviations of 51.91 and 55.81. Clearly the 6 data points with low values have invalidated the classical estimates. It is to be noted that when the data are pooled in the second quantile plot and the GD fit used, the intercept estimates the mean at 466.3 and the slope estimates the standard deviation at 21.52. With 6 low values and one high value deleted from the pooled data, the mean is 469.04 and the standard deviation is 17.00. The GD mean was near the "cleaned data" mean and its SD not too much higher. If one uses GD on the "cleaned data" the intercept which estimates the oval distance is 469.42 and the slope which estimates the SD is 17.54. Note that these are near the ordinary mean and SD for the "cleaned data." Also the normal quantile plot reveals normal like data and the GD line is followed closely by the data. This is the third graph below.

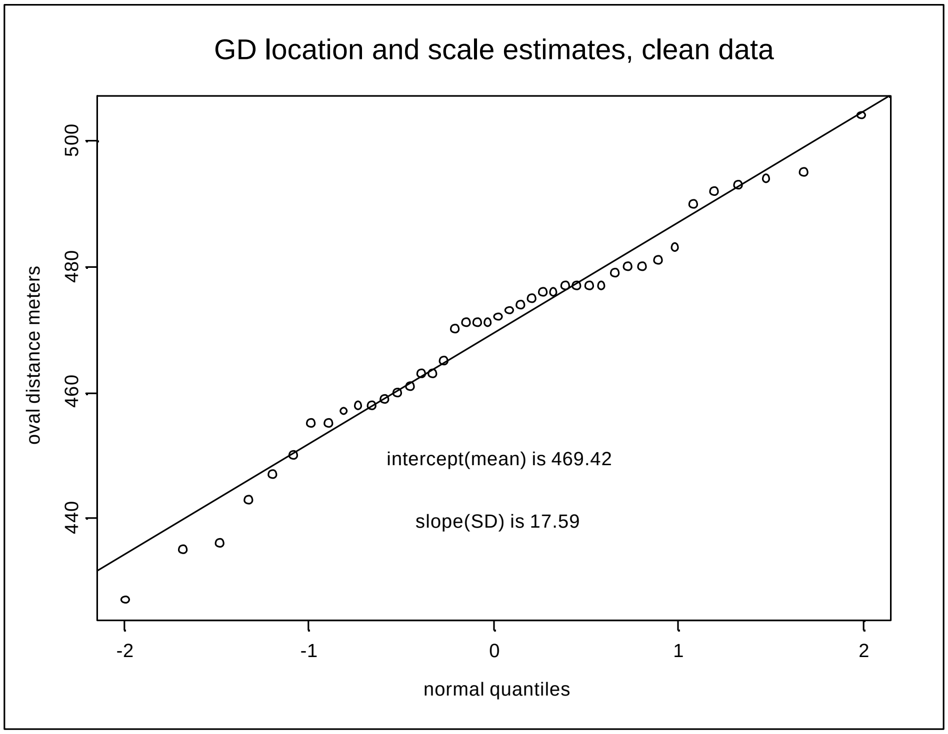
With the reduced data there are now 42 points used. The classical mean is 469.04 and its standard error is $17.00/\sqrt{42}$ = 2.62. The bootstrap technique was used to estimate the standard errors of the GD method on the mean (intercept) and the standard deviation (slope). A total of 2500 bootstrap samples were used in each case. For the mean the bias was 0.104 and the SE was 2.69 with the average bootstrap GD mean of 469.5; nearly the same as the classical method. For the estimate of standard deviation the bias was –0.871 and the standard bootstrap SE for this was 2.55. The average GD estimate of the SD was 16.72.

To show the robustness of the GD method, the same bootstrap analysis was applied to all the data, n= 49. First the regular mean is 453.57 with a SD of 53.80 so the SE is 7.69. For the GD method the observed mean is 466.3 and the bootstrap mean is 465.2 so that the bias is –1.09 and the SE is 4.15. For the bootstrap GD estimate of the SD the observed value was 21.52 with the bootstrap mean of 22.18 so that the bias was 0.67 and a SE of 4.40. The robustness of GD methods can clearly be seen as there is not the huge gap between the estimates from all the data to the “cleaned data.” In addition the bootstrap method shows that the tied value problem when handled by the maximum-minimum is not a problem for the rank based GD correlation statistic.


GD fits on oval data, o =timer, x =count
intercept and slope
timer, 471.5, 24.94
count, 459.41, 26.68
oval distance meters
250
300
350
400
450
500
-1
0
1
normal quantiles

a


GD fits on combined oval data
intercept = 466.3,slope = 21.52
oval distance meters
250
300
350
400
450
500
-2
-1
0
1
2
normal quantiles



There were students in this class who had GIS aerial photos which allowed an estimate of the distance via computer imaging methods. One student got 466 meters and the other 472. The 466 estimate was supposedly from one-meter increments while the other from a little cruder polygonal path method. It is usually unclear how many of the possible outliers to delete and in this example it is clear that without deletion the usual method is not accurate. The GD robust method however, gave good results with and without deletion.

For the two-sample problem, the first GD fit on the separate data sets reveal similar slopes (variation the same) and the difference between the intercepts (means) of (471.50 – 459.41) = 12.09. From the pooled data on the second graph the SD is estimated as 21.52 and from the bootstrap the SE as 4.15. The “standard score” is 12.09/4.15 = 2.91. This indicates a possible difference between the two methods. A classical two-sample t-test assuming homogeneous variance for all 49 points has a P-value of 0.346 and the same test with the “cleaned data” of 42 data, 7 points removed, gave a P-value of 0.084. Both the Kolmogorov-Smirnov and the Mann-Whitney test were less significant. However, when only 6 data points were removed the two-sample t-test gave a P-value of less than 0.05. So it is not exactly clear if both timer and count method are estimating the same parameter. What is clear is that the GD method gives additional information about the data analysis.

248.00	248.00	.00	1
355.00	267.00	1.00	2
427.00	317.00	1.00	3
450.00	355.00	.00	4
457.00	397.00	1.00	5
458.00	408.00	1.00	6
459.00	427.00	.00	7
460.00	435.00	1.00	8
461.00	436.00	1.00	9
463.00	443.00	1.00	10
465.00	447.00	1.00	11
471.00	450.00	.00	12
474.00	455.00	1.00	13
476.00	455.00	1.00	14
477.00	457.00	.00	15
480.00	458.00	.00	16
483.00	458.00	1.00	17
490.00	459.00	.00	18
492.00	460.00	.00	19
493.00	461.00	.00	20
494.00	463.00	.00	21
495.00	463.00	1.00	22
504.00	465.00	.00	23
533.00	470.00	1.00	24
267.00	471.00	.00	25
317.00	471.00	1.00	26
397.00	471.00	1.00	27
408.00	472.00	1.00	28
435.00	473.00	1.00	29
436.00	474.00	.00	30
443.00	475.00	1.00	31
447.00	476.00	.00	32
455.00	476.00	1.00	33
455.00	477.00	.00	34
458.00	477.00	1.00	35
463.00	477.00	1.00	36
470.00	477.00	1.00	37
471.00	479.00	1.00	38
471.00	480.00	.00	39
472.00	480.00	1.00	40
473.00	481.00	1.00	41
475.00	483.00	.00	42
476.00	490.00	.00	43

477.00	492.00	.00	44
477.00	493.00	.00	45
477.00	494.00	.00	46
479.00	495.00	.00	47
480.00	504.00	.00	48
481.00	533.00	.00	49

Five possible titles for this analysis

Summary of an Analysis of Data Measuring the Distance in Meters around the UM Oval

An Illustration of the Greatest Deviation Correlation Coefficient's (GD) Robustness in the Two-Sample Problem

Use of the Bootstrap method and Permutation Tests with a Nonparametric, rank based, correlation coefficient on Location and Scale Problems

An Example of a Subset of a General Method of Statistical Analysis with correlation coefficient's. For further Analysis see the following WEB site which contains papers on this General Method. www.math.umt.edu/gideon and email: gideon@mso.umt.edu

Exemplification of the Classical Analysis Confusion

Here is a quick summary of the 49 observations to measure the distance around the UM oval by a timer, n_t=24, or step counting, n_c=25, method. There are possibly 4-7 usual measurements(4 for the count and 3 for the timer). The null hypothesis is that the count and timer methods are equivalent, the independent two-sample problem. A sub hypothesis is that the scale factors are equal.

1. For all the data: GD analysis on location differences has p-value; of 0.063 by a permutation test, whereas the classical p-value is 0.218. For scale differences the GD permutation test p-value is 0.410 and the classical F-test gives a 0.363 p-value.
2. For the "clean" data, n=42: GD permutation test p-value is 0.076 and the classical t-test p-value is 0.085. For scale the GD permutation test p-value is 0.276 and the classical F-test p-value is 0.153.

Estimation of location and scale parameters via the bootstrap for GD compared to the classical methods.

All	Count		Timer		Combined			
	loc	scale	loc	scale	loc	SE	scale	SE
GD	459.4	26.7	471.5	24.9	466.3	4.2	21.5	4.4
Classic	446.4	51.9	461.0	55.8	453.6	7.7	53.8	11.1

Clean	Count		Timer		Combined			
	loc	scale	loc	scale	loc	SE	scale	SE
GD	466.1	17.0	473.3	20.1	469.4	2.7	17.6	2.4
Classic	465.3	14.7	472.8	18.6	469.0	2.7	17.0	1.8

The GD method is actually a general correlation method. Let q be the ordered quantiles of a N(0,1) random variable. Correlation r could be GD or any correlation coefficient. Let y^0 be the ordered y data and then solving $r(q, y^0 - bq) = 0$ for b yields an estimate of scale. The location estimate is then the median $(y^0 - bq)$.

The BCA , bias corrected accelerated method for a confidence interval, is now constructed for the distance around the oval. The method used is outlined on pages 262-

265 in J.J. Higgins' book, Introduction to Modern Nonparametric Statistics. First the bootstrap distribution of the estimate of the distance is constructed in Splus, bootstrap(data, rgrgc(normal quantiles, data, 0)$intercept) , data are the 42 ordered student measurement, and normal quantiles come from qnorm(1:42/43). From the "replicates" output comes the 1000 observations of the bootstrap distribution. From this comes the estimate of bias, z_0. Then the Jackknife estimates are used to correct for any skewness. Splus is used in computing 42 estimates, rgrgc(42 normal quantiles[-i],ordered data[-i],0)$intercept where [-i] is the Splus notation to delete the ith obervation, i=1,2,…,42. From these numbers the sum of squares and cubes are used to compute the skewness number a. It turned out that a = 0.0100 and z_0 = -0.00501. This lead to the 2.67% and 97.66% points of the bootstrap distribution as the lower and upper bounds for the 95% confidence interval. The interval was 464.06 to 474.60. The length of this BCA confidence interval is 10.54. If the bootstrap Standard Error, 2.69, of the estimate 469.4 is used to form a confidence interval $469.4 \pm 1.96(2.69)$ exactly the same length interval is obtained, 10.54.