NORMS, ORDER STATISTICS, AND VARIATION ESTIMATES

Classically a norm in statistics is essentially the same as a norm in general mathematical analysis. In this work a norm is developed in a completely different, but much more general way, namely via the correlation coefficient.

The author's previous work on location and scale estimates from correlation coefficients will be combined to produce a generalized alternative to the classical norm, called an order norm, as it is based on order statistics. This norm does agree with the classical norm on certain regular data vectors, but this new norm, in contrast to the classical norm, is robust on the unchanged data. Many current robust methods begin with data adjustments to eliminate outlier influence. This method requires no such manipulation.

This paper develops the order norm, shows it is robust for a particular correlation coefficient and that it agrees with the classical norm on certain symmetric data. It illustrates several important properties of the norm and it is shown how to produce a new inner product, a new covariance, and yet another correlation coefficient which leads to further avenues of research. An elaborate example on a classification problem using satellite data is given. The illustrations use the Greatest Deviation correlation coefficient because this nonparametric correlation coefficient makes apparent the generality of the method and gives a robust norm. Any of the correlation coefficients discussed in Gideon (G0, 2000) could be subjected to the same treatment and their particular properties discussed.

In order to develop the material, the connection between the usual least squares estimate of variation, s^2, and the order statistic least squares estimate of s is explained.

1. Preliminaries: Notation.

Random Variable Z with cumulative distribution function F(z).

R.V. X with distribution function $F\left(\frac{x-m}{s}\right)$ of the continuous kind.

The standardized R.V. $Z=\frac{X-m}{s}$, and the random sample $X_1, X_2, \cdots, X_n$.

The order statistics $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ and the column vector $x^o=(x_{(1)}, x_{(2)}, \cdots, x_{(n)})'$.

The order statistic transformation $X_{(i)}=m+sZ_{(i)}$ and expectation $E(X_{(i)})=m+sE(Z_{(i)})$, i=1,2,…n. In order to condense notation let k denote a set of constants that are sometimes centered symmetrically about zero. For now let $k_i=E(Z_{(i)}), i=1,2,\cdots n.$ When a norm based on order statistics is introduced a different k will be defined.

Before proceeding we need a result contained in Randles and Wolfe problem 1.2.4. Introduction to the Theory of Nonparametrics.

Let $D(m,s) = \sum (x_{(i)} - m - sk_i)^2$. Then it is easy to show from least squares that the minimum of D with respect to m and s is $\hat{m}_o = \bar{x}$ and $\hat{s}_o = \frac{\sum k_i x_{(i)}}{\sum k_i^2}$ where $\sum_{i=1}^{n} k_i = 0$.

It follows that $E(\hat{m}_o) = m$ and $E(\hat{s}_o) = s$. The subscript o signifies an order statistic estimate.

These estimates of location and scale were shown to be fully efficient for a normal population by Lloyd (1952). This idea is discussed in David (1970) in the location and scale chapter. None of these authors looked at the estimation through the use of correlation coefficients as is done in this paper.

Let r_p represent Pearson's correlation coefficient, then with vector notation

$$r_p(\underline{x}, \underline{y}) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

. Let $k' = (k_1, k_2, \cdots, k_n)$ where k_i is the expectation of the i[th] order statistic of the standardized random variable. In what follows $\bar{k} = 0$ because $\sum k_i = 0$, and then with k as an argument in r_p the correlation formula simplifies.

Proposition 0: $\sum_{i=1}^{n} k_i = 0$

Proof: By definition, $\sum_{i=1}^{n} k_i = \sum \left(\frac{E(X_{(i)}) - m}{s} \right) = \frac{E(\sum X_{(i)}) - nm}{s} = \frac{E(\sum X_i) - nm}{s} =$

$\frac{\sum E(X_i) - nm}{s} = \frac{nm - nm}{s} = 0.$

2. Relationships between two least squares estimates of variability

Proposition 1: Pearson's correlation coefficient gives the order statistic estimate $\hat{s}_o$.

This formulation is important because other correlation coefficients can be used in the same manner to obtain an estimate of s . Let r_p be Pearson's correlation coefficient, then because $\sum k_i = 0$, $r_p(\underline{k}, \underline{x^o} - \tilde{s}\underline{k}) = 0$ implies $\tilde{s} = \frac{\sum k_i x_{(i)}}{\sum k_i^2} = s \frac{\sum k_i z_{(i)}}{\sum k_i^2} = \hat{s}_o$.

Proposition 2: $s_{LS}^2 = \hat{s}_o^2 \frac{\sum k_i^2}{n-1} + \frac{SSE}{n-1}$ where

$$SSE = \sum (x_{(i)} - \bar{x} - \hat{s}_o k_i)^2 = s^2 \left[\sum (z_{(i)} - \bar{z})^2 - \frac{(\sum k_i z_{(i)})^2}{\sum k_i^2} \right]$$

Proof: Let us carry out the least squares regression of $\underline{x}^o$ on $\underline{k}$ or in model form $x_{(i)} = m + sk_i + e_i, i = 1,2,\cdots,n.$ First note that because $E(X_{(i)}) = m + sk_i$, $E(e_i) = 0$ which is the usual regression assumption. Of course, however, the epsilons are not independent and identically distributed. The least squares regression estimates were given by the minimum of $D(m, s)$ above. So we use the regression notation SSX = SSR + SSE, total corrected sum of squares equals regression sum of squares plus error sum of squares. One figure with four parts is given to illustrate this regression. The data $\underline{x}^o$ and $(-\underline{x})^o$ are each regressed on $\underline{k}$, normal quantiles. The data is N(0,1). Parts c and d of the Figure 1 lets kk= (-2.5,-1.5,-0.5,0.5,1.5,2.5) and regresses $\underline{x}^o$ and $(-\underline{x})^o$ on kk. The fits given are from the Greatest Deviation correlation coefficient, and least squares. This figure also helps in the understanding of later Propositions. Table 0 gives details of the fits along with the Pearson correlation coefficient fit; that is, least squares. The location and scale parts of the order norm as defined later are given as well as the Norm and the to be defined Order Norm This is the regression with kk as the horizontal axis variable. The Proposition 2 statement is for Least Squares or the Pearson fit.

Table 0: Statistics for the regression plots

	$\underline{x}^o$ *on k*		$(-\underline{x})^0$ *on k*		$\underline{x}^o$ on *kk*		$(-\underline{x})^o$ *on kk*	
	int	slope	int	slope	int	slope	int	slope
GD	0.2034	0.3932	-0.2034	0.3932	0.2052	0.1700	-0.2052	0.1700
LS or P	0.2622	0.5840	-0.2622	0.5840	0.2622	0.2803	-0.2622	0.2803
For this x the norm is 1.39151 and the Order Norms are 0.87108(GD), 1.3369(P)								

For completeness the x vector is given to five places: (-0.46387,-0.0084136,0.13033, 0.27001,0.45019,1.19520)

SSX = $\sum(x_{(i)} - \bar{x})^2 = \sum(x_i - \bar{x})^2$ and so SSX/(n-1) = s^2_{LS} is the usual least square estimate of s^2.

Now the regression line values are $\hat{x}_{(i)} = \bar{x} + \hat{s}_o k_i$ and so SSR = $\sum(\hat{x}_{(i)} - \bar{x})^2 = \hat{s}^2_o \sum k_i^2 = \hat{s}_o \sum k_i x_{(i)}$.

Finally, SSE = $\sum(x_{(i)} - \hat{x}_{(i)})^2 = \sum(x_{(i)} - \hat{m} - \hat{s}k_i)^2 = \sum(x_{(i)} - \bar{x} - \hat{s}_o k_i)^2$. Thus,

$$\frac{SSX}{n-1} = \frac{SSR}{n-1} + \frac{SSE}{n-1} \text{ gives } s^2_{LS} = \hat{s}^2_o \frac{\sum k_i^2}{n-1} + \frac{SSE}{n-1}.$$

For the $\underline{x}^o$ on kk Pearson regression in Table 0, $\sum k_i^2 = 17.5$ and SSX = SSR + SSE or $\sum(x_i - \bar{x})^2 = \hat{s}^2_o \sum k_i^2 + \sum(x_{(i)} - \hat{x}_{(i)})^2$ and with the data 1.52369 = $(0.28031)^2$ *17.5 + 0.14865 = 1.37504 + 0.14865 and dividing by 5, $\hat{s}^2_{LS}$ = 0.30474 = 0.27501 + 0.02973.

For the $\underline{x}^o$ on expected values of the order statistics with the approximation used in Table 1 the $\sum k_i^2 = 4.1941$ and for SSX = SSR + SSE we have 1.52369 = $(0.58402)^2$ *4.1941 + 0.09317 = 1.4305 + 0.09317.

Finally for this data the least squares estimate of s is $\sqrt{0.30474} = 0.55203$ while the order statistics estimates from Table 0 are 0.5840(LS) and 0.3932(GD).

Proposition 3: If $x_{(i)} = m + sk_i, i = 1,2,\cdots,n$; that is, all $e_i = 0$

$$s_{Ls}^2 = \hat{s}_o^2 \frac{\sum k_i^2}{n-1} = s^2 \frac{\sum k_i^2}{n-1}$$

Proof: If $x_{(i)} = m + sk_i, i = 1,2,\cdots,n$, then $\hat{s}_o = \frac{\sum k_i (m + sk_i)}{\sum k_i^2} = s$, and

$SSE = \sum (m + sk_i - m - sk_i)^2 = 0$, so that $s_{LS}^2 = \hat{s}_o^2 \frac{\sum k_i^2}{n-1} + 0 = s^2 \frac{\sum k_i^2}{n-1}$.

This Proposition will be used when a norm is introduced to show some norm properties.

It is instructive to have $\sum_{i=1}^{n} k_i^2$ for some values of n for the standardized normal distribution. From H.A. David (Order Statistics, page 65) or Blom, G. (1958) there is a well known approximation for $k_i = E(Z_{(i)}) \cong \Phi^{-1}(\frac{i-(3/8)}{n+(1/4)})$, where Φ is the cdf of a N(0,1) R.V.

Table 1	n	2	3	4	5	10
approx	$\sum k_i^2$	0.6949	1.5117	2.380	3.27	7.95
true		0.636	1.4323	2.300	3.195	
	n	20	40	50	75	100
approx		17.63	37.30	47.20	72.01	96.87
true						
	n	200	500	1000	5000	10000
approx		196.54	496.11	995.77	4995.008	9994.675
true						

Because $E(\sum_{i=1}^{n} Z_i^2) = E(\sum_{i=1}^{n} Z_{(i)}^2) = n$ for a standardized normal random variable, it is easy to show using variances that $\sum (E(Z_{(i)}))^2 < n$. It is apparent that when normal data is very good ($x_{(i)}$ near $m + sk_i$) then s_{LS}^2 underestimates s^2. This is because $\frac{\sum k_i^2}{n-1}$ is slightly less than one; e.g. at n = 20, $s_{LS}^2 = s^2 \frac{17.63}{19} = 0.92 s^2$ while $\hat{s}_o = s$.

3. Heuristic Motivations

Other Geometrical and Analytical Motivations for the Order Statistic Estimation of Variation

The least squares and order statistic estimates are related analytically as follows.

Note that $s_{Ls}^{2} = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum (x_{(i)} - \bar{x})^2}{n-1} = \frac{\sum (x_{(i)} - \bar{x})(m + s z_{(i)} - \bar{x})}{n-1} =$
$\frac{s \sum (x_{(i)} - \bar{x}) z_{(i)}}{n-1}$.

Because $E(s_{LS}^{2}) = s^2$, $E(\frac{\sum (x_{(i)} - \bar{x}) z_{(i)}}{n-1}) = s$. Now the order statistic estimate is obtained by replacing $z_{(i)}$ by $E(z_{(i)}) = k_i, i = 1,2,\cdots,n$. So we have

$$s_{LS}^{2} \cong \frac{s \sum (x_{(i)} - \bar{x}) k_i}{n-1} = \frac{s \sum (x_{(i)} - \bar{x}) k_i}{(n-1) \sum k_i^2} \sum k_i^2 = s \hat{s}_o \frac{\sum k_i^2}{n-1}.$$

We notice that $E(s \hat{s}_o \frac{\sum k_i^2}{n-1}) = s^2 \frac{\sum k_i^2}{n-1}$. So replacing $z_{(i)}$ by k_i for all i has resulted in biasing the estimate of s^2 but $E(\hat{s}_o) = s$.

A geometric derivation looks at SSX, the sum of squares of the vertical distances of the $x_{(i)}$ from the $\bar{x}$ line, and SSE, the sum of squares of the vertical distances of the $x_{(i)}$ from the order statistic regression line whose slope is $\hat{s}_o$. These lines are shown in Figure 1. For each point $(x_{(i)}, k_i)$, the regression line, and the $\bar{x}$ line let q be the angle between them at the horizontal origin. Then it is approximately true that

$$\tan q = slope = \hat{s}_o \cong \frac{x_{(i)} - \bar{x}}{k_i}.$$

Then, $\hat{s}_o k_i \cong x_{(i)} - \bar{x}$ and for least squares $\hat{s}_o^2 k_i^2 \cong (x_{(i)} - \bar{x})^2$ so that

$\hat{s}_o^2 \frac{\sum k_i^2}{n-1} \cong \frac{\sum (x_{(i)} - \bar{x})^2}{n-1} = s_{LS}^{2}$. This approximation should be valid for other techniques for fitting a regression line to estimate s. For good data the different estimates should be close together, but for bad data s_{LS}^{2} will be, in general , an inflated estimate of variability and the other correlation techniques should be more valid. These same statements hold for the order statistics norm that is introduced.

4. An Order Statistic Norm (or Partial Norm)

Let r be any correlation coefficient and solve for $\hat{s}_o$ in the equation $r(\underline{k}, \underline{x}^o - \hat{s}_o \underline{k}) = 0$ and let $\hat{m}_o$ be either the mean or median of the uncentered residuals $(\underline{x}^o - \hat{s}_o \underline{k})$. This method is discussed in Gideon(scale estimation). Use of the mean would tend to make the norm less robust. So for nonparametric correlation coefficient norms the median is used.

Recall that the usual square of the Euclidean norm is $\|\underline{x}\|^2 = \underline{x}' \underline{x} = n\bar{x}^2 + \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Definition 1: The Order Statistic Norm

$\|\underline{x}\|_o^2 = n\hat{m}_o^2 + \hat{s}_o^2 \sum_{i=1}^{n} k_i^2$.

In Table 0, two order norms are given, one with GD and one with Pearson's correlation coefficient. The intercept of the kk regression is $\hat{m}_o$, and n = 6.

In Proposition 3, if there is a strict linear relationship between the $x_{(i)}$ and the k_i then these two norms will be the same if $\hat{s}_o = s$, the slope of the linear relationship.

Example 1: Let r be Pearson's r and from earlier $\frac{\sum k_i x_{(i)}}{\sum k_i^2} = \hat{s}_o$. Then the mean of the uncentered residuals is $\bar{x}$ and $\|\underline{x}\|_o^2 = n\hat{m}_o^2 + \hat{s}_o^2 \sum_{i=1}^{n} k_i^2 = n\bar{x}^2 + \left(\frac{\sum k_i x_{(i)}}{\sum k_i^2} \right)^2 \sum k_i^2$. In order to make the order norm comparable to the usual norm we now take $k_i = (i - \frac{n+1}{2}), i = 1,2,\cdots,n$; that is, the first n integers centered at 0. Because now $\sum k_i^2 = \frac{n(n^2-1)}{12}$ this order statistic norm with Pearson's correlation coefficient is $\|\underline{x}\|_o^2 = n\bar{x}^2 + \frac{12}{n(n^2-1)} (\sum (i - \frac{n+1}{2}) x_{(i)})^2$. Call this the Pearson order norm. In certain cases, the order statistic norm should agree with the regular norm. One case is when n equals two. When n = 2, $\|\underline{x}\|^2 = 2\bar{x} + \frac{1}{2}(x_1 - x_2)^2$ and for the Pearson order norm

$\|\underline{x}\|_o^2 = 2\bar{x}^2 + \frac{12}{2*3} (\frac{-1}{2} x_{(1)} + \frac{1}{2} x_{(2)})^2 = 2\bar{x}^2 + \frac{1}{2}(x_{(2)} - x_{(1)})^2 = \|\underline{x}\|^2$.

All other correlation coefficient norms should also be equivalent to the usual norm when n = 2 because there is no error in fitting a slope to two points. However, for nonparametric correlation coefficients (based on ranks) the computation depends on the tie breaking procedure introduced in Gideon and Hollister (1987). The idea is simple but necessary. Let r be any nonparametric correlation coefficient , say Kendall's tau or the Greatest Deviation as in G&H. Let $(\underline{x}, \underline{y})$ be the data in which some or even all values could be tied. Then let $r^+(\underline{x}, \underline{y})$ be the correlation coefficient value when ties are broken to produce ranks that give the largest possible value. Similarly let $r^-(\underline{x}, \underline{y})$ be the least possible value. Then r is defined to be the average of these two extremes. The first use of this will be when all y values are the same and when x and y are converted to ranks, the x vector for r^+ will be 1,2,3,…,n while the y vector in ranks will all be same value. In this scenario r^+ will be one and r^- will be minus one so that r =0. Let r_{NP} represent a nonparametric correlation coefficient.

Proposition 4: For an r_{NP} order norm where $\hat{m}_o = median(\underline{x}^o - \hat{s}_o \underline{k})$ with n = 2, $\|\underline{x}\|^2 = \|\underline{x}\|_o^2$.

Proof: For n =2, $k_1 = \frac{-1}{2}$ and $k_2 = \frac{+1}{2}$ and the slope of the line is $x_{(2)} - x_{(1)}$. It is now demonstrated that with $\hat{s}_o = x_{(2)} - x_{(1)}$, $r_{NP}(\underline{k}, \underline{x}^o - \hat{s}_o \underline{k}) = 0$. We have $\underline{x}^o - \hat{s}_o \underline{k} = \begin{pmatrix} x_{(1)} - \hat{s}_o(-1/2) \\ x_{(2)} - \hat{s}_o(+1/2) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x_{(1)} + x_{(2)} \\ x_{(1)} + x_{(2)} \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \bar{x} \end{pmatrix}$. Thus, $\hat{m}_o = \bar{x}$. When the k vector is converted to ranks it is (1,2) while the residual vector has two tied values. Thus $r_{NP}(\underline{k}, \underline{x}^o - \hat{s}_o \underline{k}) = 0$ because it is the average of +1 and –1 which is zero. So we have $\|\underline{x}\|_o^2 = 2\bar{x}^2 + (x_{(2)} - x_{(1)})^2(\frac{1}{4} + \frac{1}{4}) = 2\bar{x}^2 + \frac{1}{2}(x_{(1)} - x_{(2)})^2 = \|\underline{x}\|^2$. This is an application of Proposition 3 with $\sum (x_i - \bar{x})^2 = \hat{s}_o^2 \sum k_i^2$ where $m = \frac{x_{(1)} + x_{(2)}}{2}, \hat{s}_o = s = x_{(2)} - x_{(1)}$

Example 2: If $\underline{x} = (1,2,\cdots,n)'$ and as before $k_i = i - \frac{n+1}{2}$ for all i, then $\|\underline{x}\|_o^2 = \|\underline{x}\|^2$.
Proof: This again illustrates the use of Proposition 3. The relationship between the points

$$(k_i, x_{(i)}) = (i - \frac{n+1}{2}, i)$$

is strictly linear with a slope of $s = 1$. Now $\underline{x}^o - s\,\underline{k} = \frac{n+1}{2}\underline{1}$ where $\underline{1}$ is a vector of 1’s. In the solution of $r_{NP}(\underline{k}, \underline{x}^o - 1 * \underline{k}) = 0$ the second argument has all ties and hence by the G&H tied value procedure $\hat{s}_o = s = 1$ (as in the above n=2 example, the correlation coefficient is the average of +1 and –1 which is zero). Also the median of the set of all tied values is the mean of the first n positive integers, $\frac{n+1}{2}$. Also $\sum k_i^2 = \sum (i - \frac{n+1}{2})^2 = \sum (x_i - \bar{x})^2$. Hence, we have $\|\underline{x}\|_o^2 = n\bar{x}^2 + 1 * \sum (i - \frac{n+1}{2})^2 = \|\underline{x}\|^2$. It is known that $\sum (i - \frac{n+1}{2})^2 = \frac{n(n^2 - 1)}{12}$.

5. Properties of the Order Statistic Norm

As stated in Corollary 1.3.33 of Randles and Wolfe, if the vector of observations is the outcome of a random sample of a random variable X, which is symmetric about point μ, the sample mean, $\bar{x}$, is an odd translation statistic, SSX is a even translation-invariant statistic, and $\text{cov}(\bar{x}, SSX) = 0$. The same properties hold for the parts of the Order Statistic Norm, $\hat{m}_o$ and $\hat{s}_o$. In our proofs we use the standard properties of correlation coefficients and symmetry, and the results hold for all such standard correlation coefficients

Proposition 5: $\hat{s}_o$ is an even translation-invariant statistic

First the invariant part follows from the translation invariance property of correlation coefficients. Let $y_j = x_j + c,\ j = 1,2\cdots n$. If $\hat{s}_o$ is the solution of $r_{NP}(\underline{k}, \underline{x}^o - \hat{s}_o \underline{k}) = 0$, then

$r_{NP}(\underline{k}, \underline{y}^o - \hat{s}_o \underline{k}) = r_{NP}(\underline{k}, (\underline{x+c})^o - \hat{s}_o \underline{k}) = r_{NP}(\underline{k}, \underline{x}^o + c - \hat{s}_o \underline{k}) = r_{NP}(\underline{k}, \underline{x}^o - \hat{s}_o \underline{k}) = 0$.

For the evenness we need to show that $\hat{s}_o(x) = \hat{s}_o(-x)$. Let $x_{(1)} < \cdots x_{(i)} < \cdots < x_{(n)}$ and $-x_{(1)} > \cdots > -x_{(i)} > \cdots > -x_{(n)}$ or in terms of order statistics of –x $(-x)_{(n)} > \cdots > (-x)_{(n+i-1)} > \cdots > (-x)_{(1)}$. That is, $(-x)_{(i)} = -x_{(n+1-i)} i = 1,2,\cdots,n$. We also use the symmetry of the k vector, $k_i = -k_{n+1-i}, i = 1,2,\cdots,n$. In vector notation this becomes $-\underline{x}^o =$ the transpose of $\left((-x)_{(n)}, \cdots, (-x)_{(n+1-i)}, \cdots, (-x)_{(1)}\right)$. The reverse of this column vector contains the order statistics for –x and, letting $rev(\underline{k})$ be the k vector in reverse order, we have $\underline{k} = -rev(\underline{k}) = rev(-\underline{k})$. Now $r_{NP}(\underline{k}, \underline{x}^o - \hat{s}_o \underline{k}) = 0$ and so $-r_{NP}(\underline{k}, \underline{x}^o - \hat{s}_o \underline{k}) = r_{NP}(-\underline{k}, -\underline{x}^o - \hat{s}_o(-\underline{k})) = r_{NP}(\underline{k}, (-\underline{x})^o - \hat{s}_o \underline{k}) = 0$, upon reversing the order of the x and k vectors for this last equality to be true. Thus, $\hat{s}_o(x) = \hat{s}_o(-x)$, and the scale operation is even.

Proposition 6: $\hat{m}_o$ is an odd translation statistic.

First the translation property. If $\underline{y} = \underline{x} + c$ then from Proposition 5, $\hat{s}_o$, for both the x and y, is the same. Let $\hat{m}_o$ be the mean or median of the uncentered residuals $(\underline{x}^o - \hat{s}_o \underline{k})$. Then $\hat{m}_o(y) = median(\underline{x}^o + c - \hat{s}_o \underline{k}) = c + median(\underline{x}^o - \hat{s}_o \underline{k})$ =c+ $\hat{m}_o$. The same is true if the mean rather than the median is used. For the odd property, the same notation is used as in Proposition 5, and either the mean or median can be used and again the statistic $\hat{s}_o$ stays the same. If $x_{(i)} - \hat{s}_o k_i < 0$ then $-x_{(i)} + \hat{s}_o k_i > 0$, but this can be written as $(-x)_{(n+1-i)} - \hat{s}_o k_{n+1-i} > 0$. This holds for all i and a similar result holds fi the residuals are positive. This means that $median(\underline{x}^o - \hat{s}_o \underline{k}) = -median((-\underline{x})^o - \hat{s}_o \underline{k})$ or $\hat{m}_o(\underline{x}) = -\hat{m}_o(-\underline{x})$.

Proposition 7: For the GDCC $\|-\underline{x}\|_o = \|\underline{x}\|_o$

The proof follows from the previous two propositions.

$$\|-\underline{x}\|_o^2 = n(\hat{m}_o(-x))^2 + (\hat{s}_o(-x))^2 \sum k_i^2 = n(\hat{m}_o(x))^2 + (\hat{s}_o(x))^2 \sum k_i^2 = \|\underline{x}\|_o^2$$

Proposition 8: $\hat{m}_o$ and $\hat{s}_o$ are uncorrelated when the data are random

We would like the order statistic norm to have nearly the same properties as the usual Euclidean norm. By Proposition 5, $\hat{s}_o$ is even translation-invariant and by Proposition 6, $\hat{m}_o$ is an odd translation statistic and so by Corollory 1.3.33 of Randles and Wolfe $\text{cov}(\hat{m}_o, \hat{s}_o) = 0$ which is the same property for $\bar{x}$ and SSX. In vector notation

$\bar{x}\underline{1}'(\underline{x}-\bar{x}\underline{1})=0$ is the equivalent of the cosine or covariance of $\bar{x}$ and SSX being zero. The comparable form for the order norm is $r(\hat{m}_o\underline{k}, \underline{x}^o - \hat{s}_o\underline{k}) = 0$.

Example 3: if $\underline{x}=(c,c,\cdots,c)'$, then $\|\underline{x}\|^2 = \|\underline{x}\|_o^2$

It is clear that $\|\underline{x}\|^2 = nc^2$. For the order norm $\hat{s}_o = 0$ and hence, $\hat{m}_o = median(\underline{x}^o) = c$. Thus, $\|\underline{x}\|_o^2 = n\hat{m}_o^2 + 0 = nc^2$.

Example 4: If $\underline{y} = c\underline{x}$, then $\|\underline{y}\|_o^2 = c^2\|\underline{x}\|_o^2$. For now let c>0.

By the properties of the scale and location statistics $\hat{s}_o(y) = c\hat{s}_o(x)$ and $\hat{m}_o(y) = c\hat{m}_o(x)$. Therefore, $\|\underline{y}\|_o^2 = n\hat{m}_o{}^2(y) + \hat{s}_o^2(y)\sum k_i^2 = nc^2\hat{m}_o^2(x) + c^2\hat{s}_o^2(x)\sum k_i^2 = c^2\|\underline{x}\|_o^2$. Proposition 7 can be used to obtain the result when c < 0.

Example 5: If $\underline{y} = \underline{x} + c$, then $\|\underline{y}\|^2$ and $\|y\|_o^2$ have comparable forms

By the scale and location properties $\hat{s}_o(y) = \hat{s}_o(x)$ and $\hat{m}_o(y) = \hat{m}_o(x) + c$. Thus, we have $\|\underline{y}\|_o^2 = n(\hat{m}_o(x)+c)^2 + \hat{s}_o^2(x)\sum k_i^2$. The form for the usual norm is $\|\underline{x}+c\underline{1}\|^2 = n(\bar{x}+c)^2 + \sum(x_i - \bar{x})^2$. This is a comparable form.

6. The Order Inner Product, Covariance, and Correlation Coefficient

From the order norm it is possible to define an inner product, a covariance, and a corresponding correlation coefficient. In the following expression V is variance, V(X+Y)=V(X) + V(Y) +2COV(X,Y). Let the equivalent sample sum of squares for these terms be SSX, SSY, SS(X+Y), and SSXY where for example, SSXY = $\sum(x_i - \bar{x})(y_i - \bar{y})$ and SS(X+Y) = $\sum(x_i - \bar{x} + y_i - \bar{y})^2$. Then the sample version of the formula is SS(X+Y)=SSX + SSY +2SSXY. Now let superscript o denote the order statistic equivalent; SS^o(X+Y) = $\hat{s}_{o,x+y}^2 \sum k_i^2$, etc. Then to define the order statistic sample covariance function we need to compute from their part of the order norm function, SS^o(X+Y), SS^oX, and SS^oY, and make the analogy with the standard sample covariance function.

Definition 2: The Order Covariance function

SS^oXY = (SS^o(X+Y) – SS^oX –SS^oY)/2

Because SSXY = $\sum x_i y_i - n\overline{xy}$ where in inner product notation $<\underline{x},\underline{y}>=\sum x_i y_i$ it is now possible to define the order "inner product" $<\underline{x},\underline{y}>_o$. First recall that $<\underline{x},\underline{y}>=(SS(X+Y)-SSX-SSY+2n\overline{xy})/2$.

Definition 3: The Order Inner Product is defined as
$<\underline{x},\underline{y}>_o=(SS^o(X+Y)-SS^oX-SS^oY+2n\hat{m}_{o,x}\hat{m}_{o,y})/2$

Definition 4: The Order Correlation Coefficient

$$\hat{r}_o=\frac{SS^oXY}{\sqrt{(SS^oX)(SS^oY)}}=\frac{\hat{s}^2_{o,x+y`}-\hat{s}^2_{o,x}-\hat{s}^2_{o,y}}{2\hat{s}_{o,x}\hat{s}_{o,y}}$$

In this definition it is probability best to use for the k vector the expected values of the order statistics so that unbiased estimate of the standard deviations are obtained.

Example 6: $<\underline{x},\underline{x}>_o=\|\underline{x}\|^2_o$

This is easy to show by substitution and the property that $\hat{s}^2_{o,2x}=4\hat{s}^2_{o,x}$; the equivalent of V(2X) = 4V(X).

In these next few examples, it is shown that the order correlation coefficient has some regular features of correlation coefficients.
Example 7: If X = Y $\hat{r}_o=1$

Any legitimate correlation coefficient must have this property and the property in example 8. This property follows because $\hat{s}_{o,x}=\hat{s}_{o,y}$ and $\hat{s}^2_{o,x+y}=4\hat{s}^2_{o,x}$.
Example 8: If Y = -X, $\hat{r}_o=-1$.

Now because X+Y=0, from definition 4 $\hat{r}_o=\dfrac{\hat{s}^2_{o,0}-\hat{s}^2_{o,x}-\hat{s}^2_{o,-x}}{2\hat{s}_{o,x}\hat{s}_{o,-x}}=\dfrac{0-2\hat{s}^2_{o,x}}{2\hat{s}^2_{o,x}}=-1$

Example 9: If X and Y are independent, the order correlation coefficient, $\hat{r}_o$, is estimating zero.
If the correct expectations of the standardized order statistics are used for the vector k and because V(X+Y) = V(X) + V(Y), then $\hat{s}_{o,x+y}$ estimates $\sqrt{s^2_x+s^2_y}$ in an unbiased fashion, see G3. Then the numerator of $\hat{r}_o$ is estimating $s^2_x+s^2_y-s^2_x-s^2_y$=0.

Example 10: The Order correlation coefficient with Pearson's r

Because Pearson's correlation coefficient has a closed form solution for the Order correlation coefficient it can be explicitly stated. Let $(x+y)_{(i)}$ be the i^{th} order statistic for the sum of the x and y data. Then from Section 1 and Definition 4 for Pearson's r

$$\hat{r}_o = \frac{(\sum k_i (x+y)_{(i)})^2 - (\sum k_i x_{(i)})^2 - (\sum k_i y_{(i)})^2}{2\sum k_i x_{(i)} \sum k_i y_{(i)}}.$$

In this formula the terms $\sum k_i^2$ have cancelled.

7. What is a Zero for the Order Norm?

The Order Norm has been tested on some real data, but before that can be discussed it must be understood what constitutes a zero when the Order Norm uses the Greatest Deviation correlation coefficient in its definition. Other correlation coefficients would have somewhat different zeros and other properties. Table 2, given below, shows how much of the data must be zeros to have $\|\underline{x}\|_o = 0$. The general rule is that if about 2/3^{rds} of the data are zeros then the GDCC Order Norm is zero. This means that 1/3rd of the most extreme part of the data does not have an influence on the Norm. Thus, the Order Norm is robust when up to 1/3rd of the data is erratic. When almost all of the data points are good these order methods and classical methods are approximately equally effective.

Table 2	
Sufficient number of zero coordinates for the Greatest Deviation Order Norm to be Zero	
Sample Size	Number of Zeros needed to produce a Zero
7	6
14	10
21	14
36	25
96	65
600	401

This table was produced by an actual computer run of the Order norm with the GDCC as its defining criteria. Recall that a zero means that both $\hat{m}_o$ and $\hat{s}_o$ are zero. This happens because the GDCC is a point counting scheme that is solved for the variation estimate, and also the median is used for the location.

Example 11: The Triangle Inequality does not hold for the r_{gd} Order norm.

For a 7 dimensional vector let $\underline{x}_i$ be a vector with an i in the ith position and zeros elsewhere. Then from Table 2, all the Order norm values of the $\underline{x}_i$ are zero, but the Order norms of the sums are not zero. Thus, the Order norm of the sum is not less than or equal to the Order norm of the individual vectors. Let $\underline{y}_j = \sum_{i=1}^{j} \underline{x}_i,\ j = 1,2,\cdots,7$. Table 3 shows the norms of the $\underline{y}_j$ vectors so that a comparison can be made between the Order norm and the Euclidean norm.

Table 3: Values of the Order norm and Euclidean norm on the $\underline{y}_j$ vectors			
j	$\underline{y}_j$	Order Norm	Euclidean norm
1	(1,0,0,0,0,0,0)	0	1
2	(1,2,0,0,0,0,0)	1.247	2.236
3	(1,2,3,0,0,0,0)	3.754	3.741
4	(1,2,3,4,0,0,0)	5.647	5.472
5	(1,2,3,4,5,0,0)	7.483	7.416
6	(1,2,3,4,5,6,0)	9.539	9.539
7	(1,2,3,4,5,6,7)	11.832	11.832

It is actually the case that both norms agree exactly for j = 6 and 7. Because the triangle inequality does not hold with GDCC, the Order norm is not a true norm, and perhaps it should be called a "partial" or "pseudo" norm. Computer simulations do show that this inequality seems to hold for data generated from continuous random variables. The triangle inequality would have to be investigated for each correlation coefficient to see if it holds, but the use of the median would suggest otherwise.

Example 12: Satellite data and Land cover classifications
This example comes from Satellite data where the color spectrum of the land surface and land cover type were recorded. The elevation was also recorded and put in units comparable to the light spectrum readings. Eight color bands and the elevation were recorded for polygonal regions of average size 0.09 square kilometers. Only six of the color bands were found to be useful classifiers for this data. Based on this information the land surface is to be classified into various types. This data was training data in order to construct a classification rule and then estimate the accuracy of various classification methods by cross validation. This was accomplished by classifying each of 2528 observations using the other 2527 observations. The norms of the differences were calculated. Two nearest neighbor methods were used in which the K closest points were selected and the observation being classified was assigned to the group that had the most observations among these K points. The usual K-nearest neighbor (K-NN) classification used Euclidean distance to measure closeness (the Euclidean norm).

In order to promote a better understanding of the Order norm, some standard statistics are given for two points and their difference. Because the Order norm is computed via location and scale estimates, these types of statistics are listed for both the Euclidean and Order norms. These two points and their difference are given in Table 5. The norms of the differences were used in the nearest neighbor selection process. In Table 4 these two seven-dimensional points are labeled x_1 and x_2. These points are from Group 4212 and the Order norm for K = 10 assigned them to the correct group while the Euclidean norm did not. The classical statistics include the median, mean, standard deviation and the corresponding norm. The Order norm statistics include the location estimate and the analogous standard deviation. Since $\frac{\sum k_i^2}{n-1} = \frac{n(n^2-1)}{12(n-1)} = \frac{n(n+1)}{12}$ and for n = 7 this is $\frac{14}{3}$, the comparable estimate to the classical standard deviation from

Section 3 is $\hat{s}_o\sqrt{\frac{14}{3}}$ where $\hat{s}_o$ is the slope of the line in the Order norm development. Label this SD_o.

Table 4: Statistics for two observations and their difference								
	Classical statistics				Order method statistics			
	median	mean	SD	norm	$\hat{m}_o$	$\hat{s}_o$	SD_o	O norm
x_1	57	63.43	40.59	195.06	57.75	14.75	31.86	171.57
x_2	60	68.14	39.75	204.90	64.08	15.92	34.39	189.31
x_1-x_2	-3	-4.71	5.02	17.52	-4.33	1.66	3.60	14.46

The actual values used above are given in the table below.

Table 5	TM1	TM2	TM3	TM4	TM5	TM7	Elev
x_1	57	24	26	43	66	140	88
x_2	60	26	29	49	80	138	95
x_1-x_2	-3	-2	-3	-6	-14	2	-7

Other than the Order norm being slightly less than the classical norm, no obvious reason appears as to why the Order norm and not the classical norm correctly classified these two points. If more points are looked at, the Order norm is usually less than the classical norm using GDCC because of its robustness. The location estimate from the Order norm most often is between the classical mean and the median as it was for the above data.

Table 6: The number of correct classifications for Landcover Type (Group) Nearest Neighbor classification when K = 10								
Group	3100	3300	4101	4203	4205	4206	4212	4223
n	277	657	18	207	52	108	442	143
Norm	166	569	0	136	17	93	317	55
O norm	126	572	0	128	9	68	342	52
Group	4230	4260	4270	4300	6100	7800	Totals	
n	24	38	112	25	281	144	2528	
Norm	2	11	51	2	247	98	1764	
O norm	1	6	50	3	228	82	1667	

In examining Table 6 for the effectiveness of the Order norm it is seen that the Order norm is better for Group 4212 and nearly the same for Groups 3300, 4101, 4203, 4223, 4230, 4270, 4300 and possibly not as good in the remaining six groups. The overall effectiveness of the Order norm was 0.659 whereas it was 0.698 for the classical method. As the “true” classification is necessarily subjective it is not really clear which method is best. The closeness of the two methods in many groups and one being better in other groups demonstrates, however, that the Order method does allow a different and valuable robust look at the classification procedure. Group #4212 is the Douglas Fir land cover type. This type is a closed-canopy forest which is pure, or nearly pure Douglas Fir. Because co-occurring species (e.g. Ponderosa Pine) are common, the type is easily confused with several others (e.g. Douglas Fir/Ponderosa Pine, Douglas Fir/Lodgepole Pine).

8. Summary
The development of an alternative norm has promise for any data that might not be perfect or cannot be screened. With the use of the Order norm with GDCC it is sufficient to use the vector components as they are and hence, avoid the difficult and subjective task of deleting or re-weighting suspicious data. The use of the Order norm on the Satellite data showed that except for the Douglas Fir land surface types the relative low classification ability of the Euclidean norm is not due to the influence of outliers. Satellite data is such that massive amounts are gathered and elaborate screening of the data seems impractical. Mathematical and statistical techniques are needed that are robust and straight forward. Only one nonparametric correlation coefficient was tested on real data in this paper but other correlation coefficients could be used to construct Order norms and then tested for their properties.

BIBLIOGRAPHY

David, H.A. (1970), *Order Statistics,* Wiley & Sons, Inc.

Lloyd, E.H. (1952), *Least Squares Estimation of Location and Scale Parameters Using Order Statistics*, Biometrika, (39), p88-95.

(G0)Gideon, R.A. A Generalized Interpretation of Pearson's r.

(G1)Gideon, R.A. and Hollister, R.A. (1987), A Rank Correlation Coefficient Resistant to Outliers, Journal of The American Statistical Association, Vol. 94, 14,656-666.

(G2)Gideon, R.A., Rummel, S. A., and Li, H. (1994), Nonparametric Correlation and Regression Methods. Unpublished

(G3)Gideon, R.A., Li, H., Rummel, S.A., Bruder, J., Lee, L.C., and Thiel, M. (1992). Robust Location and Scale Estimation with Nonparametric Correlation Coefficients, unpublished.

(G4)Gideon, R.A., Rummel, S., and Li, Hongzhe (1993), Multiple Linear Regression Using Nonparametric Correlation Coefficient r_g. Unpublished

(G5)Gideon,RA.,Prentice,M.J.,and Pyke,R.(1989). The Limiting Distribution of the Rank Correlation Coefficient r_g, Appears in Contributions to Probability and Statistics (Essays in Honor of Ingram Olkin) edited by Gleser,L.,J., Perlman,M.D.,Press,S.J., and Sampson,A.R. Springer-Verlang ,N.Y. pp 217-226.

(R)Rummel, S.E. (1991). A Procedure for Obtaining a Robust Regression Employing the Greatest Deviation Correlation Coefficient. Unpublished Ph. D.