A Simple Noncalculus Proof That the Median Minimizes the Sum of the Absolute Deviations

NEIL C. SCHWERTMAN, A. J. GILKS, and J. CAMERON*

It is widely known among statisticians that the median minimizes the sum of the absolute deviation about any point for a set of x's, $x_1, x_2, x_3, \ldots, x_n$. Some authors (e.g., David 1970) point out that when n is even the minimizing point is not necessarily unique. The proof that the median minimizes the sum of the absolute deviations is omitted in many mathematical statistics textbooks. Other textbooks (e.g., Bickel and Dobsum 1977, p. 54; Cramer 1946, p. 179; De Groot 1975, p. 170; Dwass 1970, p. 341; Von Mises 1964, pp. 373–374) suggest or prove the result using expectation and integral calculus for continuous data. Wasan (1970, p. 119) used a similar expectation argument for discrete distributions only. Sposito, Smith, and McCormick (1978) provided a somewhat involved proof using summations. Bloomfield and Steiger (1983) provided a rather difficult and more general investigation of the minimization of the general L_p norm that, when $p = 1$, proves the median minimizes the sum of the absolute derivations. A somewhat simplified calculus proof was given by Shad (1969). Aitken (1952, pp. 32–34) provided a clever proof with only minor use of calculus but did not provide a convenient computational procedure. Some authors, such as Gentle, Sposito, and Kennedy (1977), have used the absolute deviation for various applications, whereas the general L_p norm has been used by others. For example, Sielken and Hartley (1973) and Sposito (1982) showed how one can obtain unbiased L_1 and L_p estimators.

First consider some difficulties that must be overcome when using differentiation in the proof. From any set of x's construct the ordered set $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, where $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$, and define the sum of the absolute deviations about any point, say α, as $D(\alpha) = \Sigma_{i=1}^{n} |x_{(i)} - \alpha|$. A proof using differentiation of $D(\alpha)$ requires considerable care, since $D(\alpha)$ is nondifferentiable at $\alpha = x_{(i)}$ ($i = 1, 2, \ldots, n$). In addition, $D(\alpha) = \Sigma_{i=1}^{k} (\alpha - x_{(i)}) + \Sigma_{i=k+1}^{n} (x_{(i)} - \alpha) = (2k - n)\alpha - \Sigma_{i=1}^{k} x_{(i)} + \Sigma_{i=k+1}^{n} x_{(i)}$, where $x_{(k)} \le \alpha \le x_{(k+1)}$. The fact that k and each of the two summations are functions of α must be considered in all differentiations.

The following noncalculus proof is a simple alternative, based on sets, that is appropriate for continuous or discrete populations, readily demonstrates the nonuniqueness for even n, and provides a convenient method of computation. Consequently, this proof should be easier for most students to understand.

Theorem. For any set of n finite x's in R^1, the sum of the absolute deviations about a point α is minimized when the point is the median.

Proof. First, consider any two x's, $x_2 > x_1$. Then for any point α such that $x_1 \le \alpha \le x_2$ the sum of the absolute deviations about α is $\alpha - x_1 + x_2 - \alpha = x_2 - x_1$. For $\alpha \notin [x_1, x_2]$ and if $\alpha < x_1$, however, the sum of the absolute deviations is $x_1 - \alpha + x_2 - \alpha = x_1 + x_2 - 2\alpha > x_1 + x_2 - 2x_1 = x_2 - x_1$, and if $\alpha > x_2$ the sum of the absolute deviations is $\alpha - x_1 + \alpha - x_2 = 2\alpha - x_1 - x_2 > 2x_2 - x_1 - x_2 = x_2 - x_1$. Therefore, for any two x's the sum of the absolute deviations about a point α is minimized when $\alpha \in [x_1, x_2]$ and is equal to $x_2 - x_1$.

Now consider the set of nested intervals $[x_{(1)}, x_{(n)}]$, $[x_{(2)}, x_{(n-1)}]$, $\ldots$, $[x_{(i)}, x_{(n+1-i)}]$, where $x_{(1)} \le \cdots \le x_{(n)}$, and $i = 1, 2, \ldots, c$, where c is equal to $n/2$ if n is even and is equal to $(n + 1)/2$ if n is odd. Note that when n is odd the innermost interval is $[x_{((n+1)/2)}, x_{((n+1)/2)}]$ or equals the point $x_{((n+1)/2)}$ (median). If we choose any point, say α, such that $\alpha \in \cap_{i=1}^{c} [x_{(i)}, x_{(n+1-i)}]$, then α minimizes the sum of the absolute deviations from the endpoints of each of the nested intervals. But $D(\alpha) = \Sigma_{j=1}^{n} |x_{(j)} - \alpha| = (|x_{(1)} - \alpha| + |x_{(n)} - \alpha|) + (|x_{(2)} - \alpha| + |x_{(n-1)} - \alpha|) + \cdots + (|x_{(c)} - \alpha| + |x_{(n+1-c)} - \alpha|)$. The parentheses are used to indicate the absolute deviations from the endpoints for one of the nested intervals; for example, $(|x_{(j)} - \alpha| + |x_{(n+1-j)} - \alpha|)$ is the sum of absolute deviations for the interval $[x_{(j)}, x_{(n+1-j)}]$. Since point α is in each interval, the sum within each set of parentheses or interval is minimized and hence the total sum of absolute deviation, $D(\alpha)$, is also minimized.

Note that when n is even and $x_{(n/2)} < x_{(n/2+1)}$ the innermost interval is $[x_{(n/2)}, x_{(n/2+1)}]$, and hence any α such that $x_{(n/2)} \le \alpha \le x_{(n/2+1)}$ is contained in each of the nested intervals and hence minimizes $D(\alpha)$. Also observe that for each of the nested intervals, for example, $[x_{(j)}, x_{(n+1-j)}]$, the sum of the absolute deviations of α from the endpoints is $x_{(n+1-j)} - x_{(j)}$ if α is in the interval. Then $D(\alpha) = (|x_{(1)} - \alpha| + |x_{(n)} - \alpha|) + (|x_{(2)} - \alpha| + |x_{(n-1)} - \alpha|) + \cdots + (|x_{(c)} - \alpha| + |x_{(n+1-c)} - \alpha|) = (x_{(n)} - x_{(1)}) + (x_{(n-1)} - x_{(2)}) + \cdots + (x_{(n+1-c)} - x_{(c)}) = \Sigma_{j=n+1-c}^{n} x_{(j)} - \Sigma_{j=1}^{c} x_{(j)}$.

Although the fact that the median minimizes the sum of the absolute deviations about any point is not new, this proof seems to be new or at least was not found in the literature that was reviewed. This proof is easy to illustrate graphically and should make the proof understandable to statistics students without requiring strong calculus backgrounds and without having to assume either discrete or continuous distributions.

[*Received June 1988. Revised June 1989.*]

*Neil C. Schwertman is Professor, Department of Mathematics and Statistics, California State University, Chico, CA 95929. A. J. Gilks is Senior Lecturer and J. Cameron is Senior Tutor, Division of Computing and Mathematics, Deakin University, Victoria 3217, Australia.

REFERENCES

Aitken, A. C. (1952), *Statistical Mathematics* (7th ed.), Edinburgh: Glover & Boyd.

Bickel, P. J., and Doksum, K. A. (1977), *Mathematical Statistics: Basic*

Ideas and Selected Topics, San Francisco: Holden-Day.
Bloomfield, P., and Steiger, W. L. (1983), *Least Absolute Deviations, Theory, Applications and Algorithms*, Boston: Birkhauser.
Cramer, H. (1946), *Mathematical Methods of Statistics*, Princeton, NJ: Princeton University Press.
David, H. A. (1970), *Order Statistics*, New York: John Wiley.
De Groot, M. (1975), *Probability and Statistics*, Reading, MA: Addison-Wesley.
Dwass, M. (1970), *Probability and Statistics: An Undergraduate Course*, New York: W. A. Benjamin.
Gentle, J. E., Sposito, V. A., and Kennedy, W. J. (1977), "On Some Properties of L_1 Estimators," *Mathematical Programming*, 12, 139–140.
Shad, S. M. (1969), "On the Minimum Property of the First Absolute Moment," *The American Statistician*, 23, 27.
Sielken, R. L., and Hartley, H. O. (1973), "Two Linear Programming Algorithms for Unbiased Estimation of Linear Models," *Journal of the American Statistical Association*, 68, 639–641.
Sposito, V. A. (1982), "On Unbiased L_p Regression Estimators," *Journal of the American Statistical Association*, 77, 652–653.
Sposito, V. A., Smith, W., and McCormick, G. F. (1978), *Minimizing the Sum of Absolute Deviations*, Gottingen, West Germany: Vandenhoeck & Ruprecht.
Von Mises, B. (1964), *Mathematical Theory and Probability and Statistics*, New York: Academic Press.
Wasan, M. (1970), *Parametric Estimation*, New York: McGraw-Hill.

A Simplified Derivation of the Variance of Kendall's Rank Correlation Coefficient

PAUL D. VALZ and A. IAN McLEOD*

In this article, a short and simple derivation of the variance of Kendall's rank correlation coefficient, $\hat{\tau}$, is presented by making use of the inversion vector.

KEY WORDS: Indicator function; Inversion vector.

1. INTRODUCTION

Let R_1 and R_2 be the rankings of n individuals with respect to two criteria and assume, initially, that there are no ties in either ranking. Then, without loss of generality, it may also be assumed that R_2 is in its natural order so that $R_2 = (1, 2, \ldots, n)$. Let $R_1 = (r_1, r_2, \ldots, r_n)$. Then the negative score, Q, is given by

$$Q = \sum_{i>j} I_{(0,\infty)}(r_j - r_i), \tag{1}$$

where $I_{(0,\infty)}(\cdot)$ denotes the indicator function on $(0, \infty)$. Kendall's rank correlation coefficient (Kendall 1975, eq. 1.5) is then given by

$$\hat{\tau} = 1 - 4Q/n(n - 1). \tag{2}$$

A rather lengthy derivation of the variance of $\hat{\tau}$ was given by Kendall (1975, chap. 5) for the general case of tied ranks. Noether (1967, chap. 10) presented a more concise approach. For the case in which the two criteria are assumed to be independent and continuous, however, the derivation given in Section 2 is more direct than these other approaches.

The notion of an inversion vector provides the basis for our derivation. Reingold, Nievergelt, and Deo (1977) defined an inversion vector, $I_k = (i_1, i_2, \ldots, i_k)$, as follows.

Let $X = (x_1, x_2, \ldots, x_k)$ be a sequence of numbers. A pair (x_l, x_j) is called an inversion of X if $l < j$ and $x_l > x_j$. The inversion vector of X is the sequence of integers $i_1, i_2, \ldots, i_k$ obtained by letting i_j be the number of x_l such that (x_l, x_j) is an inversion. Hence i_j is the number of elements greater than x_j and to its left in the sequence. Note that $0 \le i_j \le j - 1$. For example, the inversion vector for the permutation $P = (4, 3, 5, 2, 1, 7, 8, 6, 9)$ is $I = (0, 1, 0, 3, 4, 0, 0, 2, 0)$. It may be proved by induction that each inversion vector uniquely represents a permutation of the first k natural numbers.

2. DERIVATION OF THE VARIANCE

Let I_n be the inversion vector corresponding to the ranking R_1 so that

$$I_n = (0, i_2, i_3, \ldots, i_n), \qquad 0 \le i_j \le j - 1.$$

It follows from the definitions of Q and I_n that

$$Q = \sum_{j=1}^{n} i_j. \tag{3}$$

Under the assumption of independent rankings, inversion vectors are equiprobable. Since the set of $n!$ inversion vectors may be divided into $(n!/j)$ subsets of j inversion vectors so that members of the same subset differ only on the jth element, it then follows that each of the j possible values $(0, 1, \ldots, j - 1)$ of i_j have probability j^{-1}. Hence

$$E(i_j) = (j - 1)/2 \tag{4}$$

and, consequently,

$$E(Q) = \sum_{j=1}^{n} E(i_j) = \frac{1}{2}\sum_{j=1}^{n} (j - 1) = \frac{1}{2}\binom{n}{2}. \tag{5}$$

*Paul D. Valz is a Ph.D student and A. Ian McLeod is Associate Professor, Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario N6A 5B9, Canada. The authors acknowledge helpful comments from a referee who drew attention to Noether (1967). Support for this research was provided by the National Sciences and Engineering Research Council of Canada.