Inequalities Among Some Measures of Location

A. Laradji and Anwar H. Joarder
Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
Emails: alaradji@kfupm.edu.sa, anwarj@kfupm.edu.sa

ABSTRACT

Inequalities involving some sample means and order statistics are established. An upper bound of the absolute difference between the sample mean and median is also derived. Interesting inequalities among the sample mean and the median are obtained for cases when all the observations have the same sign. Some other algebraic inequalities are derived by taking expected values of the sample results and then applying them to some continuous distributions. It is also proved that the mean of a nonnegative continuous random variable is at least as large as \( p \) times \( 100(1 - p) \)th percentile.

Key Words and Phrases: Sample means, order statistics, inequalities

Mathematics Subject Classification: Primary 26D99; Secondary 60E15, 62G30

1. INTRODUCTION

Inequalities involving measures of location namely, sample means, median and extreme observations do not appear to be generally known. This note is inspired by Shiffler and Harsha (1980) and Macleod and Henderson (1984) who worked on the bounds of sample standard deviation. Some inequalities involving sample means and linear combinations of order statistics namely, median and extremes are established.

We believe that the inequalities will, in particular, provide additional information to students in statistics, and, in general, open a new direction of further research to refine inequalities on other sample statistics along the line of Shiffler and Harsha (1980), Macleod and Henderson (1984) and Eisenhauer (1993). Another motivation for the current research is the improved inference in situations when the parameter is known to have a restricted sample space (Ahmed, 1991). For a number of applications in econometrics and design of experiments, see Silvapulle and Sen (2004). In Section 3, it is shown how we can have restricted parameter space.

Let \( x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \) be order statistics corresponding to the sample \( \tilde{x} = (x_1, x_2, \cdots, x_n) \) with median \( \tilde{x} \) given by \( 2\tilde{x} = x_{(\lfloor n/2 \rfloor + 1)} + x_{(n/2 + 1)} \) where \( \lfloor m \rfloor \) is the bracket function denoting the largest integer not exceeding \( m \). Also let the arithmetic, geometric and harmonic means of a sample \( \tilde{x} \) be denoted by \( a(\tilde{x}) = \tilde{x} \), \( g(\tilde{x}) \) and \( h(\tilde{x}) \).
respectively. In this paper we establish interesting inequalities involving some of the sample characteristics, namely, $\bar{x}$, $g(x)$, $h(x)$, $\tilde{x}$, $x(i)$ and $x(n)$. An upper bound of the absolute difference between the sample mean and median is derived. Interesting inequalities among sample mean and median are obtained for cases when all the observations have the same sign. Some other inequalities are derived by taking expected values of the sample results and then applying them to some continuous distributions. It is also proved that the mean of a nonnegative continuous random variable is at least as large as $p$ times 100$(1 - p)^{th}$ percentile.

2. INEQUALITIES AMONG SOME MEASURES OF LOCATION AND EXTREME OBSERVATIONS

The following lemma is obvious.

**Lemma 2.1** Let $x_i \leq y_j$ $(1 \leq i \leq n)$. Then

(i) $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i$

(ii) $\prod_{i=1}^{n} x_i \leq \prod_{i=1}^{n} y_i$ if $x(i) \geq 0$.

(iii) $\sum_{i=1}^{n} \frac{1}{x_i} \geq \sum_{i=1}^{n} \frac{1}{y_j}$ if $x(i) > 0$.

Consider the three sequences $A = \{a_1, a_2, \ldots, a_{2n}\}$, $B = \{b_1, b_2, \ldots, b_{2n}\}$ and $C = \{c_1, c_2, \ldots, c_{2n}\}$ each having $2n$ terms defined by

$a_k = \begin{cases} x(i) & \text{if } 1 \leq k \leq n, \\ \bar{x} & \text{if } n + 1 \leq k \leq 2n, \end{cases}

b_k = x_{i(k/2+1/2)}$ and

$c_k = \begin{cases} \tilde{x} & \text{if } 1 \leq k \leq n, \\ x(n) & \text{if } n + 1 \leq k \leq 2n. \end{cases}$

These sequences are then

$A = \{x(i), x(i), \ldots, x(i), \tilde{x}, \tilde{x}, \ldots, \tilde{x}\}$

$B = \{x(i), x(i), x(2), x(2), \ldots, x(n), x(n)\}$ and

$C = \{\tilde{x}, \tilde{x}, \ldots, \tilde{x}, x(n), x(n), \ldots, x(n)\}$

where $A$ and $C$ contain $n$ medians ($\tilde{x}$). For $1 \leq k \leq n$,

$a_k = x(i) \leq x_{i(k/2+1/2)} = b_k \leq \frac{1}{2} (x_{i(n/2+1/2)} + x_{i(n/2+1/2)}) = \tilde{x} = c_k$ and for $n + 1 \leq k \leq 2n$,

$a_k = \tilde{x} \leq \frac{1}{2} (x_{i(n/2+1/2)} + x_{i(n/2+1/2)}) \leq x_{i(k/2+1/2)} = b_k \leq x(n) = c_k$. Since the elements of the three sets satisfy the conditions of Lemma 2.1, we have the following theorem.
Theorem 2.1 For any sample \( x_1, x_2, \ldots, x_n \) of \( n \geq 2 \) observations with \( x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \), the following inequalities hold:

(i) \( \frac{x_{(i)} + \bar{x}}{2} \leq \bar{x} \leq \frac{\bar{x} + x_{(n)}}{2} \)

(ii) \( \sqrt{x_{(i)} \bar{x}} \leq g(x) \leq \sqrt{\bar{x} x_{(n)}} \) if \( x_{(i)} \geq 0 \) and

(iii) \( \frac{2}{1/x_{(i)} + 1/\bar{x}} \leq h(x) \leq \frac{2}{1/\bar{x} + 1/x_{(n)}} \) if \( x_{(i)} > 0 \) and

where \( g(x) \) and \( h(x) \) are the geometric and harmonic means of a sample of \( n \) observations.

Proof. Applying Lemma 2.1 (i) to the sets \( A \) and \( B \), and then to \( B \) and \( C \) we have \( n x_{(i)} + n \bar{x} \leq 2n \bar{x} \) and \( 2n \bar{x} \leq nx_{(n)} + n \bar{x} \) which imply Theorem 2.1 (i). The other two parts of the theorem are deduced from Lemma 2.1 (ii) and Lemma 2.1 (iii) respectively in a similar manner.

Since in many real world situations observations are nonnegative, the following corollary may be useful.

Corollary 2.1 If \( x_{(i)} > 0 \), then \( \frac{1}{2} h(x) \leq \bar{x} \leq 2 \bar{x} \).

Proof. It follows from Theorem 2.1 (iii) that \( \frac{1}{2} h(x) \leq \frac{1}{1/\bar{x} + 1/x_{(n)}} \leq \frac{1}{1/\bar{x}} = \bar{x} \leq x_{(i)} + \bar{x} \)

which, by virtue of Theorem 2.1(i), cannot exceed \( 2 \bar{x} \).

Theorem 2.2 For any sample \( x_1, x_2, \ldots, x_n \) of \( n \geq 3 \) observations with \( x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \), the following inequalities hold:

(i) \( (i - 1)x_{(i)} + x_{(i)} + (n - i)x_{(i+1)} \leq n \bar{x} \), for \( 1 \leq i \leq n - 1 \),

(ii) \( (n - 1)x_{(i)} + (n + 1) \bar{x} \leq 2n \bar{x} \leq (n + 1) \bar{x} + (n - 1)x_{(n)} \),

(iii) \( |\bar{x} - \bar{x}| \leq \frac{n - 1}{n + 1} \max (\bar{x} - x_{(i)}, x_{(n)} - \bar{x}) \).

Proof. (i) For \( 1 \leq i \leq n - 1 \), we have

\[
n \bar{x} = \left(x_{(i)} + x_{(i+1)} + \cdots + x_{(n-1)}\right) + x_{(i)} + \left(x_{(i+1)} + \cdots + x_{(n)}\right)
\geq (i - 1)x_{(i)} + x_{(i)} + (n - i)x_{(i+1)}.
\]

(ii) For odd \( n \) and \( i = (n+1)/2 \), we have \( n \bar{x} \geq \frac{n - 1}{2} x_{(i)} + \bar{x} + \frac{n - 1}{2} \bar{x} \) from (2.1), so

that by virtue of \( \bar{x} = x_{(i+1)/2} \leq x_{(i+3)/2} \) we have

\[
2n \bar{x} \geq (n - 1)x_{(i)} + (n + 1)\bar{x}.
\]

(iii) \( |\bar{x} - \bar{x}| \leq \frac{n - 1}{n + 1} \max (\bar{x} - x_{(i)}, x_{(n)} - \bar{x}) \).
When \( n \) is even, letting \( \frac{2}{n} i = i/n \) and \( \frac{1}{2} x_i \) in (2.1), we obtain
\[
\bar{x} \left( n - 1 \right) x_i + \frac{n}{2} x_{(n/2 + 1)} \quad \text{and} \quad \bar{x} \left( n - 1 \right) x_i + \frac{n}{2} x_{(n/2 + 1)} + \left( \frac{n}{2} - 1 \right) x_{(n/2 + 2)}.
\]

By adding the above two inequalities and using the fact that \( \bar{x} \leq x_{(n/2 + 1)} \leq x_{(n/2 + 2)} \) for even \( n \), we have
\[
2n \bar{x} \geq \left( n - 1 \right) x_i + 2 \bar{x} + \frac{n}{2} \bar{x} + \left( \frac{n}{2} - 1 \right) \bar{x}
\]
so that \( 2n \bar{x} \geq \left( n - 1 \right) x_i + \left( n + 1 \right) \bar{x} \). Hence for any sample of size \( n \geq 2 \), we have
\[
\left( n - 1 \right) x_i + \left( n + 1 \right) \bar{x} \leq 2n \bar{x}.
\]

Next from \(-x_{(n)} \leq -x_{(n-1)} \leq \cdots \leq -x_{(1)} \), similarly we obtain
\[
\left( n - 1 \right) (-x_{(n)}) + \left( n + 1 \right) (-\bar{x}) \leq 2n (-\bar{x})
\]
or, \( \left( n + 1 \right) \bar{x} + \left( n - 1 \right) x_{(n)} \geq 2n \bar{x} \)
which completes the proof.

(iii) By writing \( 2n \bar{x} = \left( n - 1 \right) \bar{x} + (n + 1) \bar{x} \), it follows from (ii) that
\[
\left( n - 1 \right) x_i + \left( n + 1 \right) \bar{x} \leq \left( n - 1 \right) \bar{x} + (n + 1) \bar{x} \leq \left( n + 1 \right) \bar{x} + \left( n - 1 \right) x_{(n)},
\]
or, \( \left( n - 1 \right) \left( \bar{x} - x_{(n)} \right) \leq \left( n + 1 \right) \left( \bar{x} - \bar{x} \right) \leq \left( n - 1 \right) \left( \bar{x} - x_{(i)} \right).

It is worth noting that the inequalities \( x_{(i)} + \bar{x} \leq 2 \bar{x} \leq \bar{x} + x_{(n)} \) in Theorem 2.1 (i) can be deduced from Theorem 2.2 (ii) in the following way:
\[
n \left( x_{(i)} + \bar{x} \right) \leq n \left( x_{(i)} + \bar{x} \right) + \left( \bar{x} - x_{(i)} \right) = \left( n - 1 \right) x_{(i)} + \left( n + 1 \right) \bar{x} \leq 2n \bar{x}
\]
\[
\leq \left( n + 1 \right) \bar{x} + \left( n - 1 \right) x_{(n)} = n \left( \bar{x} + x_{(n)} \right) - \left( x_{(n)} - \bar{x} \right) \leq n \left( \bar{x} + x_{(n)} \right).
\]

**Corollary 2.2** The following inequalities hold for any sample \( x_1, x_2, \cdots, x_n \) of \( n \geq 2 \) observations:

\[
(i) \ 2 | \bar{x} | \geq | \bar{x} |, \quad \text{if the observations have the same sign.} \quad (2.6)
\]
\[
(ii) \ (n - 1)x_{(i)} \leq 2n \bar{x} - (n + 1) \bar{x} \leq (n - 1)x_{(n)}. \quad (2.7)
\]

**Proof.** (i) If \( x_{(i)} \geq 0 \), then both \( \bar{x} \) and \( \bar{x} \) are nonnegative, and \( \bar{x}/2 \leq (x_{(i)} + \bar{x})/2 \) which cannot exceed \( \bar{x} \) by Theorem 2.1(i). If \( x_{(n)} \leq 0 \) then both \( \bar{x} \) and \( \bar{x} \) are nonpositive, and by Theorem 2.1(i) we also have \( \bar{x} \leq (\bar{x} + x_{(n)})/2 \) which cannot exceed \( \bar{x}/2 \). Taking absolute values we have the inequality in (i).

(ii) The inequalities follow directly from Theorem 2.2 (ii).
Remarks

(i) If the observations \( x_1 \leq x_2 \leq \cdots \leq x_n \) have the same sign then \( 2 |\bar{x}| = |\bar{x}| \) occurs exactly when all \( x \)’s are equal to 0. If \( x_1 \geq 0 \), then \( 2 |\bar{x}| = |\bar{x}| \) implies \( 2\bar{x} = \bar{x} \) so that we have \( 0 \leq (n-1)x_1 + (n+1)\bar{x} \leq n\bar{x} \) by the left hand inequality in Theorem 2.2 (ii).
Hence, in this case, \( (n-1)x_1 + \bar{x} = 0 \) which happens only if \( \bar{x} = 0 \) i.e. if \( 2\bar{x} = 0 \) i.e. if all observations are 0’s. A similar argument applies when \( x_n \leq 0 \).

(ii) If \( x_1 > 0 \), then \( 2n\bar{x} \geq (n+1)\bar{x} \) by the left hand inequality in Theorem 2.2 (ii).
Similarly, \( 2\bar{x} < \bar{x} \) if \( x_n < 0 \).

(iii) In case not all the observations have the same sign, an example of a sample showing \( 2\bar{x} = \bar{x} \) may be: \( n = 3 \), \( x_1 = -10 \), \( x_2 = 10 \), \( x_3 = 15 \) which could be average temperatures of three days in a city.

(iv) If skewness is judged by the third central moment, a positively skewed distribution may produce a median exceeding mean and a negatively skewed distribution may produce mean exceeding median. For a brief but insightful discussion of measures of skewness, see Eisenhauer (2002).

Corollary 2.3 If \( n \geq 2 \) observations in the sample \( x_1, x_2, \ldots, x_n \) have the same sign, then \( |(\bar{x}/\bar{x}) - 1| \leq 1 \).

Proof. Since the \( x \)’s have the same sign, it follows from (2.6) that \( (\bar{x}/\bar{x}) = |\bar{x}/\bar{x}| \leq 2 \).
If \( (\bar{x}/\bar{x}) \geq 1 \), then \( |(\bar{x}/\bar{x}) - 1| = (\bar{x}/\bar{x}) - 1 \leq 1 \), and if \( (\bar{x}/\bar{x}) \leq 1 \), then \( |(\bar{x}/\bar{x}) - 1| = 1 - (\bar{x}/\bar{x}) < 1 \). Hence the proof.

3. INEQUALITIES INVOLVING EXPECTED VALUES

The following theorem follows from Corollary 2.2.

Theorem 3.1 For any nonnegative random variable \( X \), the inequality \( E(\bar{X}) \geq E(\bar{X})/2 \) holds whenever the expected values exist.

Evidently, the above holds for any symmetric distribution e.g. the uniform or the normal distribution. An example is given below for exponential distribution.

Example 3.1 Let the random variable \( X \) have the exponential probability density function (pdf)
\[
 f(x) = \beta^{-1} e^{-x/\beta} \text{ where } 0 < x, 0 < \beta.
\]
It is known that the mean and the median of the distribution are $\beta$ and $\ln 2$ respectively. An alternative but more direct proof of the inequality in Theorem 3.1 for the exponential distribution is as follows:

Since $X_1, X_2, \cdots, X_n$ are identically distributed, it follows that $E(\overline{X}) = \beta$ and hence we have to prove that $\beta \geq E(\overline{X})/2$. For $n = 2m + 1$, it is easy to check that

$$E(\overline{X}) = \left(\frac{2m + 1}{m!}\right) \beta I(m), \ I(m) = \int_0^\infty u e^{-(m+1)u} (1-e^{-u})^m du.$$ 

Thus we have to prove that

$$\frac{(2m + 1)!}{(m!)^2} \beta I(m) \leq 2\beta. \quad (3.1)$$

Since $I(m) \leq B(m, m + 1)/e$, it follows from (3.1) that $\frac{(2m + 1)!}{(m!)^2} \beta I(m) \leq 2\beta$ for all $m \geq 2$. Note that $I(1) = 5/36$. Hence for all $m \geq 1$ and $n = 2m + 1$, we obtain $E(\overline{X}) \leq \beta = E(\overline{X})$. Similarly this can be proved for $n = 2m$ (Laradji and Joarder, 2002). Alternatively, it can be quickly verified by Harter and Balakrishnan (1996, 42).

We now apply Theorem 3.1 to some continuous distributions and obtain interesting inequalities:

(i) For the above exponential distribution, the expected value of the $i$th order statistic $X_{(i)}$ is given by $E(X_{(i)}) = \beta \sum_{j=1}^i (n - j + 1)^{-1}$ (see Harter and Balakrishnan, 1996, 42), and $E(\overline{X}) = E(X) = \beta$ so that for $n = 2m + 1$, it follows from Theorem 3.1 that

$$\sum_{j=1}^m (2m - j)^{-1} \leq 2 \quad \text{while for} \quad n = 2m, \quad \text{we have} \quad \sum_{j=1}^m \left( \frac{(m + 1 - j)^{-1} + (m + 2 - j)^{-1}}{} \right) \leq 3.$$

(ii) For the gamma distribution with p.d.f.

$$f(x) = \frac{1}{\beta^n \Gamma(\alpha)} x^{\alpha - 1} e^{-x/\beta}, \quad 0 \leq x, \quad 0 < \beta, \quad 0 < \alpha,$$

the expected value of the $i$th order statistic $X_{(i)}$ is given by

$$E(X_{(i)}) = \frac{n \beta}{\Gamma(\alpha)} \left( \frac{n - 1}{\alpha} \right) \int_0^\infty \left( \frac{\Gamma(\alpha; x)}{\Gamma(\alpha)} \right)^{\alpha - 1} \left( 1 - \frac{\Gamma(\alpha; x)}{\Gamma(\alpha)} \right)^{n-i} x e^{-x} dx$$

where $\Gamma(\alpha; x) = \int_0^x t^{\alpha-1} e^{-t} \ dt$ (see Harter and Balakrishnan, 1996, 45), and

$$E(\overline{X}) = E(X) = \alpha \beta$$

so that for $n = 2m + 1$, it follows from Theorem 3.1 that

$$\int_0^\infty \left( \frac{\Gamma(\alpha; x)}{\Gamma(\alpha)} - \frac{\Gamma^2(\alpha; x)}{\Gamma^2(\alpha)} \right)^m x e^{-x} dx \leq 2\Gamma(\alpha + 1) B(m + 2, m), \quad (3.2)$$

and for $n = 2m$, we have
\[ \int_0^\infty \left( \frac{\Gamma(\alpha; x)}{\Gamma(\alpha)} - \frac{\Gamma^2(\alpha; x)}{\Gamma^2(\alpha)} \right)^{m-1} x^{\alpha} e^{-x} \, dx \leq 2\Gamma(\alpha+1) B(m,m). \]  

(3.3)

which is a slightly better bound than (3.2).

(iii) For the Weibull distribution with p.d.f.

\[ f(x) = \alpha x^{\alpha-1} e^{-x/\beta}, \quad 0 \leq x, \quad 0 < \beta, \quad 0 < \alpha, \]

the expected value of the \(i\)th order statistic \(X_{(i)}\) is given by

\[ E(X_{(i)}) = n \beta \binom{n-1}{i-1} \Gamma(1+1/k) \sum_{j=0}^{i-1} (-1)^{i-1-j} \binom{i-1}{j} (n-j)^{-1-1/k} \]

(Harter and Balakrishnan, 1996, 44), and \( E(\bar{X}) = E(X) = \beta \Gamma(1+1/k) \) so that for \( n = 2m+1 \), it follows from Theorem 3.1 that

\[ \sum_{j=0}^{m} (-1)^{m+j} \binom{m}{j} (2m+1-j)^{-1/\alpha} \leq 2B(m+1,m+1), \]

(3.4)

and for \( n = 2m \), we have

\[ \sum_{j=0}^{m} (-1)^{m+j} \binom{m-1}{j} (2m-j)^{-1/\alpha} \leq \frac{B(m,m)}{m} \left( \frac{2}{m+1} + \frac{1}{m^{2+1/k}} \right). \]

(iv) For the Pareto distribution with p.d.f.

\[ f(x) = \alpha \theta x^{\alpha-1}, \quad 0 < \theta \leq x, \quad 0 < \alpha, \]

the expected value of the \(i\)th order statistic \(X_{(i)}\) is given by

\[ E(X_{(i)}) = \frac{\Gamma(n+1) \Gamma(n-i+1-1/\alpha)}{\Gamma(n-i+1) \Gamma(n+1-1/\alpha)} \theta \]

(see Harter and Balakrishnan, 1996, 71), and \( E(\bar{X}) = E(X) = \alpha \theta (\alpha-1)^{-1}, \quad 1 < \alpha \)

(Johnson, Kotz and Balakrishnan, 1994, 577) so that for \( n = 2m+1 \), it follows from Theorem 3.1 that

\[ \frac{\Gamma(2m+2)}{\Gamma(m+2)} \frac{\Gamma(m+2-1/\alpha)}{\Gamma(2m+2-1/\alpha)} \leq \frac{2\alpha}{\alpha-1}, \]

and for \( n = 2m \) we have

\[ \frac{\Gamma(2m+1)}{\Gamma(2m+1-1/\alpha)} \left( \frac{\Gamma(m+1-1/\alpha)}{\Gamma(m+1)} + \frac{\Gamma(m-1/\alpha)}{\Gamma(m)} \right) \leq \frac{2\alpha}{\alpha-1}. \]

4. THE MEAN AND QUANTILE INEQUALITY

The following lemma is well known.

**Lemma 4.1** If \( X \) is a nonnegative random variable, then

\[ E(X) = \int_0^\infty (1-F(x)) \, dx \]
where \( F(x) \) is the cumulative distribution function (cdf) of \( X \).

**Lemma 4.2** Let \( G \) be a nonnegative decreasing function on \([0, \infty)\). Then

\[
xG(x) \leq \int_0^x G(t)\,dt.
\]

**Proof.** Since \( G(x) \leq G(t) \) for all \( 0 \leq t \leq x \), \( G(x) \) is (Riemann) integrable on each interval \([0, x]\) for \( x > 0 \) and then it follows that

\[
\int_0^x G(t)\,dt \geq \int_0^x (G(x)\,dt = xG(x) \text{ for all } x \geq 0.
\]

**Theorem 4.1** Let \( F(x) \) be the cdf of a nonnegative continuous random variable \( X \). Then \( x_p \leq \mu / p \) for each \( p \) \((0 < p < 1)\) such that \( F(x_p) = 1 - p \).

**Proof.** Since the function \( G(x) = 1 - F(x) \) is decreasing on \([0, \infty)\) and nonnegative, by Lemma 4.1 and Lemma 4.2, we have the following for all \( x \geq 0 \):

\[
\mu = \int_0^\infty G(t)\,dt \geq xG(x).
\]

Hence for each \( p \in (0, 1) \), if we denote by \( x_p \), the real number \( x \) such that \( F(x) = 1 - p \), we have \( G(x_p) = p \) and hence \( \mu \geq px_p \).

It is worth noting that in case \( p = 1/2 \), it follows from Theorem 4.1 that \( \tilde{\mu} \leq 2\mu \) where \( \tilde{\mu} = x_{0.5} \), the median of \( X \).

**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge the excellent research support provided by King Fahd University of Petroleum and Minerals, Saudi Arabia. The authors also thank an anonymous referee and Prof. N. Balakrishnan for constructive suggestions that have improved the quality of the paper.

**REFERENCES**


