King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 321

June 2004

Some Inequalities in Descriptive Statistics

A. H. JOARDER and A. LARADJI

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

Some Inequalities in Descriptive Statistics

A. H. JOARDER and A. LARADJI

Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
E-mails: anwarj@ kfupm.edu.sa, alaradji@kfupm.edu.sa

Abstract Some inequalities involving sample mean, median and standard deviation are established. These are used to find inequalities for some well known sample statistics namely, coefficient of variation, coefficient of skewness as well as the least squares estimator of the slope and intercept parameters.

1. Introduction

Page and Murty (1982, 1983) published an elementary proof of the inequality $|\tilde{x}-\mu| \le \sigma$, while O'Cinneide (1990) presented a new proof for $|\tilde{x}-\mu|\le\sigma$ and stated the following generalization:

Proposition 1.1 Let X be a random variable with mean μ and standard deviation σ, $0<p<1$ and $q=1-p$. Then the following inequality holds

$|x_{(p)}-\mu|\le\sigma \max\left(\sqrt{p/q},\ \sqrt{q/p}\right)$ where $x_{(p)}$ is the p-th quantile.

For $p=1/2$, $x_{(p)}=\tilde{x}$, the sample median, it follows from the above proposition that $|\tilde{x}-\mu|\le\sigma$. Dharmadhikari (1991) noted that for $p\ne 1/2$, the inequality is somewhat unsatisfactory. The refined inequality proved by her with the help of one-sided Chebyshev inequality is stated in the following theorem.

Proposition 1.2 Let X be a random variable with mean μ and standard deviation σ, and $0<p<1$. Then the following inequalities hold

$$\mu-\sigma\sqrt{q/p}\le x_{(p)}\le\mu+\sigma\sqrt{q/p}$$

For stimulating discussions readers may go through Mallows (1991) and the references therein. A more general inequality that relates sample standard deviation to mean and p-th order statistics discussed by David (1988) and David (1991) is given below:

Theorem 1.1 Let $x_{(p)}$ be the p th order statistics from a sample of size n. Then we have the inequality

$$| x_{(p)} - \bar{x} | \le s \ \max\left(\sqrt{\frac{(n-1)(p-1)}{n(n+1-p)}}, \sqrt{\frac{(n-1)(n-p)}{np}} \right)$$

Sample standard deviation (s) or variance (s^2) is nonnegative, and is defined by $(n-1)s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$. But for most data sets, the bounds are much narrower intervals than the nonnegative part of the real line. It may be mentioned that the sample variance has been viewed from different perspectives by Joarder (2001).

It is well known that for $n = 2$, the sample variance has a simpler form given by $s^2 = w^2 / 2$ where $w = x_{(n)} - x_{(1)}$ is the sample range with $x_{(i)}$ $(i = 1, 2, ..., n)$, the ith sample observation. Shiffler and Harsha (1980) formulated an upper bound for the sample standard deviation (s) in terms of the sample range w, while Mcleod and Henderson (1984) determined a lower bound for s in terms of w. Eisenhauer (1993) combined them. A stronger version of these results is provided in Theorem 2.1.

Theorem 1.2 (Macleod and Henderson (1984) and Shiffler and Harsha (1980)).Let w and s denote the range and standard deviation of a sample of size n . Then

$$\frac{w}{\sqrt{2\,(n-1)}} \le s \le \frac{w}{2} \sqrt{\frac{n}{n-1}} .$$

By the use of Theorems 1.1 and 1.2 we immediately obtain the following corollaries:

Corollary 1.1 Let $x_{(p)}$ be the p th order statistics from a sample of size $n \ge 2$. Then the following inequalities hold:

$$\sqrt{\frac{n}{\max(p-1, n-p)} - 1} \ \sqrt{\frac{n}{n-1}} \ | x_{(p)} - \bar{x} | \le s \le \frac{w}{2} \sqrt{\frac{n}{n-1}}$$

Corollary 1.2 (Eisenhauer, 1993) Let w and s denote the range and standard deviation of a sample of size n . Then

(a) $\frac{1}{\sqrt{2(n-1)}} \le \frac{s}{w} \le \frac{1}{2} \sqrt{\frac{n}{n-1}}$.

(b) $0 \le s / w \le 1/2$ as $n \to \infty$.

2. Some Inequalities in Descriptive Statistics

The following result is a refined version of Theorem 1.2.

Theorem 2.1 Let w and s respectively denote the range and standard deviation of a sample of size n . Then

$$\frac{w}{\sqrt{2\,(n-1)}} \le \sqrt{\frac{w^2}{2(n-1)} + \frac{(\tilde{x}-\bar{x})^2}{2}} \le s \le \sqrt{\frac{n(\bar{x}-x_{(1)})(x_{(n)}-\bar{x})}{n-1}} \le \frac{w}{2}\sqrt{\frac{n}{n-1}}\,.$$

It may be remarked here that the two sides of the rightmost inequality in the theorem are equal in case $x_{(1)} + x_{(n)} = 2\bar{x}$, otherwise we always have the sharper inequality $(\bar{x}-x_{(1)})(x_{(n)}-\bar{x}) < w/2$. The following corollary that follows from the above theorem is in agreement with the known result that for two observations $s = w/\sqrt{2}$ where $w = x_{(n)} - x_{(1)}$.

Corollary 2.1 For $n \ge 2$, $s \ge \frac{1}{2}\left(x_{n-[n/2]+1} - x_{[n/2]}\right)$

Theorem 2.2 Let $x_{(p)}$ be the p th order statistics from a sample of size $n \ge 2$. Then the following inequalities hold:

(i) $|\,x_{(p)} - \bar{x}\,| \le s\sqrt{\frac{n-1}{n}} \le s$ for each p

(ii) $|\,x_{(p)} - \bar{x}\,| \le s\,\frac{n-1}{\sqrt{n(n+1)}} \le s$ if $p = \frac{n+1}{2}$

(iii) $|\,\tilde{x} - \bar{x}\,| \le s\sqrt{\frac{n-1}{n}} \le s$

The well-known inequality in (ii) tells us that $\tilde{x} - s < \bar{x} < \tilde{x} + s$, or, $\bar{x} - s < \tilde{x} < \bar{x} + s$. That is, sample mean is within one standard deviation of median. The following corollary is obvious from Theorem1.1.

Corollary 2.2 Let $\bar{x}$ and $\tilde{x}$ be the sample mean and median based on a sample of size $n \ge 2$. Then the following inequalities hold:

$$\max\left(\sqrt{\frac{n}{n-1}}\,|\,\bar{x}-\tilde{x}\,|,\ \sqrt{\frac{w^2}{2(n-1)} + \frac{(\tilde{x}-\bar{x})^2}{2}}\right) \le s \le \frac{w}{2}\sqrt{\frac{n}{n-1}}$$

Theorem 2.3 For any sample size $n \ge 2$, the following inequalities hold

$|\,\bar{x} - \tilde{x}\,| \le s\sqrt{\frac{n}{n-1}} \le \sqrt{n}\ |\,\bar{x}\,|$ if all the observations are of the same sign.

The following corollary is obvious by virtue of Theorem 1.2 and Theorem 2.2.

Corollary 2.3 For any sample of size $n \ge 2$ the following inequalities hold:

$$\frac{1}{\sqrt{n-1}}\max\left(\frac{w}{\sqrt{2}},\ |\,\bar{x}-\tilde{x}\,|\right) \le s \le \frac{1}{\sqrt{n\text{-}1}}\min\left(\sqrt{n}\ |\,\bar{x}\,|,\ \frac{w}{2}\sqrt{\frac{n}{n-1}}\right)$$

when all the observations are nonnegative.

The following theorem is due to Laradji and Joarder (2002).

Theorem 2.4 For any sample of $n \geq 2$ observations $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, the following inequalities hold:

$$(i) \frac{1}{2}\left[\left(1+\frac{1}{n}\right)\tilde{x}+\left(1-\frac{1}{n}\right)x_{(1)}\right] \leq \bar{x} \leq \frac{1}{2}\left[\left(1+\frac{1}{n}\right)\tilde{x}+\left(1-\frac{1}{n}\right)x_{(n)}\right]$$

$$(ii) \,|\, \tilde{x}-\bar{x} \,|\, \leq \frac{n-1}{n+1} \max\,(\bar{x}-x_{(1)}, x_{(n)}-\bar{x})$$

$$(iii) \left|\frac{\tilde{x}}{\bar{x}}-1\right| \leq 1$$

3. Inequalities for some Useful Statistics

If follows from Theorem 1.2 that

$$\begin{aligned} &\frac{w}{\bar{x}\sqrt{2(n-1)}} \leq \frac{s}{\bar{x}} \leq \frac{w}{2\bar{x}}\sqrt{\frac{n}{n-1}}, \quad \text{if } \bar{x}>0 \\ &\frac{w}{2\bar{x}}\sqrt{\frac{n}{n-1}} \leq \frac{s}{\bar{x}} \leq \frac{w}{\bar{x}\sqrt{2(n-1)}}, \quad \text{if } \bar{x}<0 \end{aligned} \tag{3.1}$$

The following corollary is obvious from Theorems 2.2 and 2.3.

Corollary 3.1 If $n \geq 2$ observations are positive then

$$0 \leq \sqrt{\frac{n}{n-1}}\left|\frac{\tilde{x}}{\bar{x}}-1\right| \leq \left|\frac{\tilde{x}}{\bar{x}}-1\right| \leq \frac{s}{\bar{x}} \leq \sqrt{n} \;.$$

It may be remarked here that, by Theorem 2.2, the Coefficient of Skewness $\frac{\bar{x}-\tilde{x}}{s/3}$ varies in the interval $\left[-3\sqrt{\frac{n-1}{n}},\, 3\sqrt{\frac{n-1}{n}}\right]$ which is slightly narrower than the known interval $[-3, 3]$.

Theorem 3.1 Let $w_x = y_{(n)} - y_{(1),} w_x = x_{(n)} - x_{(1)}, s_{xy} = \sum(x-\bar{x})(y-\bar{y}), s_{xx} = s_x^2$**.** Then the regression coefficient $\hat{\beta}_1 = s_{xy} / s_{xx}$ satisfies the following inequalities:

$$(ii)\; -\frac{s_y}{s_x} \leq \hat{\beta}_1 \leq \frac{s_y}{s_x} \quad (ii)\; -\sqrt{\frac{n}{2}}\;\frac{w_y}{w_x} \leq \hat{\beta}_1 \leq \sqrt{\frac{n}{2}}\;\frac{w_y}{w_x}$$

Acknowledgements

The authors gratefully acknowledge King Fahd University of Petroleum and Minerals, Saudi Arabia for providing excellent research facilities.

References

David, H.A. (1988). General bounds and inequalities in order statistics. *Communications in Statistics – Theory and Methods*, **17**, 2119-2134.

David, H.A. (1991). Mean minus median: A comment on O'Cinneide. *American Statistician*. **45**(3), 257.

Dharmadhikari, S. (1991). Bounds on quantiles: A comment on O'Cinnneide. *American Statistician*. **45** (3), 258.

Eisenhauer, J. G. (1993). A measure of relative dispersion. *Teaching Statistics*, **15**(2), 37-39.

Joarder, A.H. (2002). On some representations of sample variance. *International Journal of Mathematical Education in Science and Technology,* 33(5), 772-784.

Laradji, A. and Joarder, A.H. (2002). Inequalities involving sample mean, sample median and extreme observations. *Technical Report*, No. 283. Department of Mathematical Sciences. King Fahd University of Petroleum and Minerals. Saudi Arabia.

Mallows, C. (1991). Another comment on O'Cinneide. *American Statistician*. 1991, **45** (3), 257.

Mcleod, A.J. and Henderson, G.R. (1984). Bounds for the sample standard deviation. *Teaching Statistics*, **6**(3), 72-76.

O'Cinneide, C. A. (1990). The mean is within one standard deviation of any median. *American Statistician*, **44** (4), 292.

Page, W. and Murty, V.N. (1983). Nearness relations among measures of central tendency and dispersion: part 1. *Two Year College mathematics Journal*. 13, 315-327.

Shiffler, R.E. and Harsha, P.D. (1980). Upper and Lower Bounds for the sample standard deviation. *Teaching Statistics*, **2**(3), 84-86.