King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 273

January 2002

The Hinge Method and the Halving Method for Sample Quatiles

Anwar H. Joarder

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

The Hinge Method and the Halving Method for Sample Quartiles

ANWAR H. JOARDER

Dept of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia 31261
Email: anwarj@kfupm.edu.sa

Summary
The Hinge Method and the Halving Method for sample quartiles have been examined in the light of a property that the quartiles divide the ordered sample observations into four number of segments leaving the same number of observations in each. It is found that the Halving Method satisfies the so called property.

Key Words and Phrases: Quartiles, remainders, modulus, hexatiles, octatiles, deciles, percentiles, quantiles.

1. Introduction

A sample quantile is a point below which some specified proportion of the values of a data set lies. The median is the 0.50 quantile because approximately half of all observations lie below this value. The name fractile for quantile is used by some authors (see Lapin, 1975, 52). Quartiles, hexatiles, octatiles, deciles, percentiles are special cases of quantiles.

One method for quartiles, called the Halving Method, is based on finding the median first and then finding the medians of the upper and lower halves of the data. Done so, roughly 25% observations remain below the lower quartile and 25% above the upper quartile. If median is excluded in the calculation of outer quartiles, the method satisfies equi-segmented property that the ranks of quantiles divide the ordered sample observations in four segments having the same number of observations (m) in each segment. The literature is full of different formulae for sample quartiles with various rounding notions of the corresponding ranks of quartiles. See for example Joarder and Firozzaman (2001) for a detailed survey and illustrations.

Since the sample size can be represented by

$$n = r \bmod f = fm + r,\ (r = 0, 1, 2, \cdots, f-1), \tag{1.1}$$

the number of observations in each of the $f \le n$ segments is given by

$$m(r) = (n-r)/f \qquad (1.2)$$

or m for short. Quantiles Q_{ir} $(i = 1, 2, \cdots, f-1;\ r = 0, 1, \cdots, f-1)$ of order f are the observations corresponding to the ranks R_{ir} $(i = 1, 2, \cdots, f-1;\ r = 0, 1, \cdots, f-1)$. The ranks R_{ir} of quantile of order f satisfies equi-segmented property if

$$(i)\ \uparrow R_{1r} \uparrow -1 = m \qquad (1.3a)$$

$$(ii)\ \uparrow R_{ir} \uparrow -[R_{i-1,r}] - 1 = m,\ i = 2, 3, \cdots, f \qquad (1.3b)$$

$$(iii)\ fm + r - [R_{f-1,r}] = m \qquad (1.3c)$$

where $[x]$ and $\uparrow x \uparrow$ are the floor function (largest integer not exceeding x) and the ceiling function (smallest integer at least as large as x) of x. The equation $(1.3a)$ states that the number of observations below the first quantile is m while the equation $(1.3c)$ states that the number of observations below the first quantile is m. The equation $(1.3b)$ states that the number of observations between two consecutive quantiles is m. Quartiles (quantiles with $f = 4$) are popularly denoted by Q_1, Q_2 and Q_3.

It appears that in the disguise of this property we have rigidly defined quantiles. Ranks for quartiles offered by the Hinge Method and the Halving Method have been provided for any sample size represented by (1.1). It remains open to generalize the Halving Method to deciles $(f = 10)$, percentiles $(f = 100)$ or any quantiles.

Students and instructors alike are curious to know why the formula for quartiles contain the quantity $n+1$. Why not n or $n-1$? It would be clear down the road that $n+1$ is the total of the ranks for the largest and smallest observations in the sample, and that the rank of the median is the average of the ranks of the observations. The Halving Method discussed in this paper demonstrates in an accessible way that the formulae proposed by Joarder and Firozzaman (2001) are based on good reasoning.

2. The Hinge Method

The ranks of the hinges discussed by Tukey (1977, p32-35) are quantities to improve upon the quartiles. We will call the method proposed by Tukey the Hinge Method for quartiles, the most popular method to find quartiles. It is based on finding the median first, and then finding the medians of upper and lower halves of the data. In the Hinge Method we will count the median in both halves if the size of the sample is odd.

(a) Ranks of quartiles for $n = 4m$

The observations have ranks $1, 2, \ldots, 2m, 2m+1, \ldots, 4m$. The rank of the median is

$R_{20} = (1+4m)/2 = 2m + 2/4$. Then by the hinge method
$R_{10} = [1+(2m+2/4)]/2 = m + 3/4$ and $R_{30} = [(2m+2/4)+4m]/2 = 3m + 1/4$.

(b) Ranks of quartiles for $n = 4m+1$

The observations have ranks $1, 2, \ldots, 2m, 2m+1, 2m+2 \ldots, 4m+1$. The rank of the median is $R_{21} = [1+(4m+1)]/2 = 2m+1$. Then by the hinge method
$R_{11} = [1+(2m+1)]/2 = m+1$ and $R_{31} = [(2m+1)+(4m+1)]/2 = 3m+1$.

(c) Ranks of quartiles for $n = 4m+2$

The observations have ranks $1, 2, \ldots, 2m, 2m+1, 2m+2 \ldots, 4m+2$. The rank of the median is $R_{22} = [1+(4m+2)]/2 = 2m+1+2/4$. Then by the hinge method
$R_{12} = [1+(2m+1+2/4)]/2 = m+1+1/4$
and $R_{32} = [(2m+1+2/4)+(4m+2)]/2 = 3m+1+3/4$.

(d) Ranks of quartiles for $n = 4m+3$

The observations have ranks $1, 2, \ldots, 2m, 2m+1, 2m+2, 2m+3, \ldots, 4m+3$. The rank of the median is $R_{23} = [1+(4m+3)]/2 = 2m+2$. Then by the hinge method
$R_{13} = [1+(2m+2)]/2 = m+1+2/4$ and $R_{33} = [(2m+2)+(4m+3)]/2 = 3m+2+2/4$.

The ranks are summarized in the following table:

	$n = 4m$	$n = 4m+1$	$n = 4m+2$	$n = 4m+3$
R_{ir}	$r = 0$	$r = 1$	$r = 2$	$r = 3$
R_{1r}	$m+3/4$	$m+1$	$m+1+1/4$	$m+1+2/4$
R_{2r}	$2m+2/4$	$2m+1$	$2m+1+2/4$	$2m+2$
R_{3r}	$3m+1/4$	$3m+1$	$3m+1+3/4$	$3m+2+2/4$

The ranks of the quartiles by the Hinge Method is summarized below:

(a) $n = 4m$

$$1,2,...,m,R_{10} = m+3/4,$$
$$m+1,m+2,...,2m,R_{20} = 2m+2/4,$$
$$2m+1,2m+2,...,3m,R_{30} = 3m+1/4,$$
$$3m+1,3m+2,...,4m$$

(*b*) $n = 4m+1$

$$1,2,...,m,R_{11} = m+1,$$
$$m+2,m+3,...,2m,\ R_{21} = 2m+1,$$
$$2m+2,2m+3,...,3m,R_{31} = 3m+1,$$
$$3m+2,3m+3,...,4m+1$$

(*c*) $n = 4m+2$

$$1,2,...,m+1,R_{12} = m+1+1/4,$$
$$m+2,m+3,...,2m+1,R_{22} = 2m+1+2/4,$$
$$2m+2,2m+3,...,3m+1,R_{32} = 3m+1+3/4,$$
$$3m+2,3m+3,...,4m+2$$

(*d*) $n = 4m+3$

$$1,2,...,m+1,R_{13} = m+1+2/4,$$
$$m+2,m+3,...,2m+1,\ R_{23} = 2m+2,$$
$$2m+3,2m+4,...,3m+2,R_{33} = 3m+2+2/4$$
$$3m+3,3m+4,...,4m+3$$

It is easy to check that equi-segmented property (1.3) is satisfied by quartiles offered by the Hinges Method only for $r = 0$.

3. The Halving Method for Sample Quartiles

It has been brought to our attention by Prof. Gerald Goodall, the editor of Teaching Statistics that the set of rules for quartiles proposed by Joarder and Firozzaman (2001) is a precise formulation of hinges discussed by Tukey (1977, p32-35). In fact if median of the whole data set is ignored in the calculation of hinges, then the two extreme hinges and median enjoy equi-segmented property. It thus resolves the difference between quartiles and hinges. We will call this method the Halving Method. We develop algebraic expressions for quartiles based on this argument and call this method the Halving Method.

Though the formulae for the median in the literature appear to be different, they all are equivalent. It is given by $Q_2 = (n+1)/2$ th observation. In case n is odd, $(n+1)/2$ will be an integer so that the median will be an observation with integer rank. If however, n is even, $(n+1)/2$ will lie between $n/2$ and $n/2+1$. Then using linear interpolation the median is given by $Q_2 = [(n/2 \text{ th} + \text{the next}) \text{ observation}]/2$.

Since the sample size can be represented by $n = r \bmod 4\, n = 4m + r,\ (r = 0,1,2,3)$, the number of observations in each of the four segments may be denoted by m. We now discuss the rank of quartiles for different sample sizes $n \geq 4$.

(a) Ranks of quartiles for $n = 4m$

The observations have ranks $1, 2, ..., 2m, 2m+1, ..., 4m$. The rank of the median is

$$R_{20} = \frac{1}{4m}(1+2+\cdots+4m) = \frac{1}{4m}\frac{4m(1+4m)}{2} = \frac{1+4m}{2} = 2m+0.5$$

which is between $2m$ and $2m+1$ so that the ranks of extreme quartiles are given by

$$R_{10} = \frac{1+2m}{2} = m+0.5,\ R_{30} = \frac{(2m+1)+4m}{2} = 3m+0.5 = (n-m)+0.5.$$

It is worth mentioning that in this case none of the quartiles has integer ranks.

(b) Ranks of quartiles for $n = 4m+1$

The observations have ranks $1, 2, ..., 2m, 2m+1, 2m+2 ..., 4m+1$. The rank of the median is

$$R_{21} = \frac{1+(4m+1)}{2} = 2m+1$$

which is between $2m$ and $2m+2$ so that the ranks of extreme quartiles are given by

$$R_{11} = \frac{1+2m}{2} = m+0.5,\ R_{31} = \frac{(2m+2)+(4m+1)}{2} = 3m+1.5 = (n-m)+0.5.$$

It is worth mentioning that in this case the median has an integer rank.

(c) Ranks of quartiles for $n = 4m+2$

The observations have ranks $1, 2, ..., 2m, 2m+1, 2m+2 ..., 4m+2$. The rank of the median is

$$R_{22} = \frac{1+(4m+2)}{2} = 2m+1.5$$

which is between $2m+1$ and $2m+2$ so that the ranks of extreme quartiles

$$R_{12} = \frac{1+(2m+1)}{2} = m+1,\ R_{32} = \frac{(2m+2)+(4m+2)}{2} = 3m+2 = (n-m).$$

It is worth mentioning that in this case the extreme quartiles have integer ranks.

(d) Ranks of quartiles for $n = 4m+3$

The observations have ranks $1, 2, ..., 2m, 2m+1, 2m+2, 2m+3, ..., 4m+3$. The rank of the median is

$$R_{23} = \frac{1+(4m+3)}{2} = 2m+2$$

which is between $2m+1$ and $2m+2$ so that the ranks of extreme quartiles are

$$R_{13} = \frac{1+(2m+1)}{2} = m+1,\ R_{33} = \frac{(2m+3)+(4m+3)}{2} = 3m+3 = (n-m).$$

The ranks of the quartiles by the Halving Method is summarized below:

(a) $n = 4m$

$1, 2, ..., m, R_{10} = m+0.5,$
$m+1, m+2, ..., 2m, R_{20} = 2m+0.5,$
$2m+1, 2m+2, ..., 3m, R_{30} = 3m+0.5,$
$3m+1, 3m+2, ..., 4m$

(b) $n = 4m+1$

$1, 2, ..., m, R_{11} = m+0.5,$
$m+1, m+2, ..., 2m,\ R_{21} = 2m+1,$
$2m+2, 2m+3, ..., 3m+1, R_{31} = 3m+1.5,$
$3m+2, 3m+3, ..., 4m+1$

(c) $n = 4m+2$

$1, 2, \ldots, m, R_{12} = m+1,$
$m+2, m+3, \ldots, 2m+1, R_{22} = 2m+1.5,$
$2m+2, 2m+3, \ldots, 3m+1, R_{32} = 3m+2,$
$3m+3, 3m+4, \ldots, 4m+2$

(d) $n = 4m+3$

$1, 2, \ldots, m, R_{13} = m+1,$
$m+2, m+3, \ldots, 2m+1,\ R_{23} = 2m+2,$
$2m+3, 2m+4, \ldots, 3m+2, R_{33} = 3m+3,$
$3m+4, 3m+5, \ldots, 4m+3$

The rank of the i th quartile with $n \geq 4$ observations are given by

$$R_{ir} = i\frac{(4m+r)+1}{4} = im + \ i(r+1)/4; \qquad (2.1)$$

where i and r are integers with $1 \leq i \leq 3$ and $0 \leq r \leq 3$. In practice one needs to divide the sample size by 4 and denote the remainder by r. The number of observations in each quarter is given by $m = (n-r)/4$. Then one may use the Halving Rule given in the following table to calculate quartiles.

	$n = 4m$	$n = 4m+1$	$n = 4m+2$	$n = 4m+3$
R_{ir}	$r = 0$	$r = 1$	$r = 2$	$r = 3$
R_{1r}	$m+0.5$	$m+0.5$	$m+1$	$m+1$
R_{2r}	$2m+0.5$	$2m+1$	$2m+1.5$	$2m+2$
R_{3r}	$3m+0.5$	$3m+1.5$	$3m+2$	$3m+3$

It is interesting to note that the remainder r here is also the number of quartiles having integer ranks. The quartiles given by the halving method is beautifully illustrated by Joarder and Firozzaman (2001).

It is easy to check that equi-segmented property (1.3) is satisfied by quartiles offered by the Halving Method for $r = 0,1,2,3$.

Acknowledgements

The author is grateful to the excellent research facilities available at King Fahd University of Petroleum and Minerals, Saudi Arabia.

References

Joarder, A.H. and Firozzaman, M. (2001). Quartiles for discrete data. *Teaching Statistics*, **23** (3), 86-89.

Lapin, L. (1975). *Statistics: Meaning and Method.* Harcourt Brace Jovanovich, Inc. New York.

Tukey, J.W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison Wesley.

File: p96a.doc