King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR297

May 2003

A Comparison and Contrast of Some Methods for Sample Quartiles

A H Joarder and R M Latif

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

A Comparison and Contrast of Some Methods for Sample Quartiles

Anwar H. Joarder and Raja M. Latif
Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia 31261
Emails: anwarj@kfupm.edu.sa, raja@kfupm.edu.sa

Abstract A remainder representation of the sample size is exploited to write out the ranks of quartiles exhaustively which in turn help compare ranks for quartiles offered by different methods available in the literature. The criterion of equisegmentation property that the number of ranks below that of the first quartile, that between the consecutive quartiles, and that above the third quartile are the same has been used to compare. Four segmentation identities are obtained for each method of quartiles which show clearly the number of observations in each of the four quarters if the observations are distinct. The Halving Method and the Remainder Method have been proposed for the calculation of sample quartiles. Quartiles provided by each of these two methods satisfy equisegmentation property if the observations are distinct. More interestingly, these two methods provide the number of quartiles having integer ranks.

Key Words and Phrases: Quartiles, remainders, modulus, quantiles.

1. Introduction

Quartiles, deciles, percentiles or more generally fractiles are uniquely determined for continuous random variables. A p^{th} quantile of a random variable X (continuous or discrete) is a value x_p such that $P(X < x_p) \le p$ and $P(X \le x_p) \ge p$. Let X be a continuous or discrete random variable with probability density function $f(x)$ and the cumulative distribution function $F(x) = P(X \le x)$. If the distribution is continuous, then $P(X < x_p) = p$ and $P(X \le x_p) = p$ because $P(X = x_p) = 0$. Therefore, for the continuous case, $F(x_p) = p$.

The quartiles $Q_1 = x_{0.25}$, $Q_2 = x_{0.50}$ and $Q_3 = x_{0.75}$ for a continuous random variable with cumulative distribution function $F(x)$ are defined by $F(x_{0.25}) = 0.25$, $F(x_{0.50}) = 0.50$ and $F(x_{0.75}) = 0.75$ respectively. Let X follow an exponential distribution with the probability density function

$$f(x) = \begin{cases} \dfrac{1}{\beta}\, e^{-x/\beta} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

with the cumulative distribution function $F(x) = 1 - e^{-x/\beta}$. Then

$1 - e^{-Q_1/\beta} = 1/4$, $1 - e^{-Q_2/\beta} = 2/4$ and $1 - e^{-Q_3/\beta} = 3/4$ so that

$Q_1 = \beta \ln(4/3)$, $Q_2 = \beta \ln 2$, $Q_3 = \beta \ln 4$.

However, for the discrete distribution, one has to use the basic definition. Consider the Binomial distribution $B(n = 4, \pi = 1/2)$. The probability mass function is given by

$$f(x) = \begin{cases} \binom{4}{x}(1/2)^4, & x = 0, 1, \cdots, 4 \\ 0 & \text{elsewhere} \end{cases}$$

Then $x_{0.25} = 1$, is the first quartile of the distribution since

$$P(X < 1) = P(X = 0) = 0.0625 \leq 0.25$$
$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.3125 \geq 0.25$$

Similarly $x_{0.50} = 2$, is the second quartile of the distribution since

$$P(X < 2) = 0.3125 \leq 0.50, \quad P(X \leq 2) = 0.6825 \geq 0.50$$

Note that the median is the same as 0.5-quantile or the 50th percentile, or the 5th decile. It is not surprising that the 60th percentile, $x_{0.6} = 2$, since $P(X < 2) = 0.3125 \leq 0.60$ and $P(X \leq 2) = 0.6825 \geq 0.60$. Similarly it can be checked that the third quartile is given by $x_{0.75} = 3$.

In case we have a sample (discrete in nature), it is, however, difficult to define quartiles. One method, called the Hinge Method, is based on finding the median first and then finding the medians of the upper and lower halves (including original median in both halves) of the data. Done so, roughly 25% observations remain below the lower quartile and 25% above the upper quartile. 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 the observations lie below this value. The name fractile for quantile is used by some authors (see Lapin, 1975, 52).

The literature is full of different formulae for sample quartiles with various rounding notions of the corresponding ranks for quartiles. We compare and contrast different methods of quartiles in the light of equisegmentation property that the number of ranks below that of the first quartile, that between the consecutive quartiles, and that above the third quartile are the same. For each method of quartiles, four segmentation identities are obtained which show clearly the number of observations in each of the four quarters if the observations are distinct. The Halving Method and the Remainder Method have been proposed for the calculation of sample quartiles. Quartiles provided by each of these two methods divide the ordered sample observations in four quarters with the same number of observations in each segment and provide the number of quartiles having integer ranks if the observations are distinct.

2. The Popular Method

There are many methods available for calculating sample quartiles in different elementary text books on statistics without any explanation. The most popular one, called Popular Method hereinafter, is described below. The rank of the $i(i = 1,2,3)$ th quartile is given by

$$i\,(n+1)/4 = l + d, \quad i = 1, 2, 3 \qquad (2.1)$$

where l is the largest integer not exceeding $i(n+1)/4$. Then the Popular Method uses the following linear interpolation formula for the calculation of sample quartiles

$$Q_i = x_{(l)} + d\,(x_{(l+1)} - x_{(l)}) = (1-d)\,x_{(l)} + d\,x_{(l+1)}, \quad (\mathrm{i} = 1, 2, 3), \qquad (2.2)$$

where $x_{(l)}$ is the l-th ordered observation (Ostle, Turner, Hicks and McElrath, 1996, 38). To write out the ranks exhaustively let us denote the sample size by the following remainder-modulus representation

$$n = r\,(\mathrm{mod}\,4) = 4m + r, \;(r = 0, 1, 2, 3), \qquad (2.3)$$

so that the number of observations in each of the $4 \le n$ segments is given by $m = (n-r)/4$. With this representation of the sample size the ranks and quartiles of a sample will be denoted respectively by R_{ir} and Q_{ir}; $i = 1, 2, 3;\; r = 0, 1, 2, 3$. Though quartiles Q_{ir}; $i = 1, 2, 3;\; r = 0, 1, 2, 3$ are usually denoted by Q_i; $i = 1, 2, 3$, we will not suppress r as it plays an important role in comparing the ranks of quarters given by different methods.

Let the number of observations in each segment be m_i $(i = 1,2,3,4)$. Then the equisegmentation property guarantees that $m_1 = m_2 = m_3 = m_4$ if the observations are distinct. In case $1 \le n \le 3$, the above formulae can also be used to calculate quartiles with $m = 0$.

It is interesting to note that though the Popular Method is not based on good mathematical reasoning, the equisegmentation property is satisfied by the quartiles provided by this method for all sample sizes except for $n = 4m + r$, $m \ge 1$, $r = 2$. For $r = 2$, the number of observations in four segments are $m, (m+1), (m+1)$ and m respectively.

Thus it is essential to modify the formulae of ranks so that the equisegmentation property is satisfied by quartiles provided by the Popular Method for any sample size. It is observed that, whenever $n = 4m + 2$, simple arithmetic rounding of ranks provided by this method would satisfy the equisegmentation property.

Example 2.1 The sizes of the police forces in the ten largest cities in the United States in 1993 (the numbers represent hundreds) published in *USA Today*, February 17, 1995 are given below:

1.7 1.9 2.0 2.8 3.9 4.7 6.2 7.6 12.1 29.3

(Bluman, 1997, 137).

We now calculate quartiles by Popular Method. Here the sample size is $n = 10 = 4(2) + 2$ so that $m = 2$ and $r = 2$. Since $r = 2$ we will denote the ranks of quartiles by R_{i2} $(i = 1, 2, 3)$. The rank of the quartiles provided by the Popular Method are (see equation 2.1)
$R_{12} = (n+1)/4 = 2.75,\ R_{22} = (n+1)/2 = 5.5,\ R_{32} = 3(n+1)/4 = 8.25$
so that by linear interpolation (see equation 2.2) the quartiles are given by

$$Q_{12} = x_{(2.75)} = (1-0.75)\,x_{(2)} + 0.75\,x_{(3)} = 0.25(1.9) + 0.75(2.0) = 1.975$$
$$Q_{22} = x_{(5.5)} = (1-0.5)\,x_{(5)} + 0.5\,x_{(6)} = 0.5(3.9) + 0.5(4.7) = 4.3$$
$$Q_{32} = x_{(8.25)} = (1-0.25)\,x_{(8)} + 0.25\,x_{(9)} = 0.75(7.6) + 0.25(12.1) = 8.725$$

To check the equisegmentation property, we show the position of the quartiles by downward arrows in the sample:

$$\begin{array}{cccccccccccccc} & & \Downarrow & & & & \Downarrow & & & & \Downarrow & & \\ 1.7 & 1.9 & & 2.0 & 2.8 & 3.9 & & 4.7 & 6.2 & 7.6 & & 12.1 & 29.3 \end{array}$$

We observe that there are $2\,(=m)$, $3(=m+1)$, $3(=m+1)$ and $2(=m)$ observations in the four segments i.e. the quartiles do not satisfy the equisegmentation property for $n = 4m + 2$.

3. A Review of the Well-known Formulae of Sample Quartiles

In this section we survey the formulae for quartiles available in the literature. We provide algebraic expressions for quartiles by all the methods available in the literature. The use of remainder allows us to figure out the decimal part of the formulae of ranks for quartiles for a particular sample of size n. Let $m = (n-r)/4,\ n = 4m + r \geq 4$, and R_{ir} be the rank of i th quartile with m observations in each segment. Then the rank of the i th quartile is given by

$$R_{ir} = i\,\frac{(4m+r)+1}{4} = im + i(r+1)/4 = im + [u_{ir}] + d_{ir}/4 \qquad (3.1)$$

where i and r are integers with $1 \leq i \leq 3$, $0 \leq r \leq 3$, $[u_{ir}]$ is the largest integer less than or equal to $u = u_{ir} = i(r+1)/4$ and $i(r+1) = d_{ir} (\mathrm{mod}\, 4)$. The quartiles can then be calculated by the simple linear interpolation as

$$Q_{ir} = (1 - d/4)\,x_{(im+[u])} + (d/4)\,x_{(im+[u]+1)} \qquad (3.2)$$

where $x_{(i)}$ is the i th ordered observation, $u = u(i,r) = i(r+1)/4$, $[u]$ is the greatest integer less than or equal to u and $d = d_{ir} = 4(u - [u])$. The above method will be called the Popular Method.

Method 1 (Popular Method)

The ranks for sample quartiles provided by the Popular Method can be written out exhaustively as (see 3.1):

$$R_{10} = m + 1/4,\ R_{20} = 2m + 2/4,\ R_{30} = 3m + 3/4$$
$$R_{11} = m + 2/4,\ R_{21} = 2m + 1,\ R_{31} = 3m + 1 + 2/4$$
$$R_{12} = m + 3/4,\ R_{22} = 2m + 1 + 2/4,\ R_{32} = 3m + 2 + 1/4$$
$$R_{13} = m + 1,\ R_{23} = 2m + 2,\ R_{33} = 3m + 3$$

Ranks are tabulated below for different sample sizes.

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

Segmentation identities are given by

$$m + 0R_{10} + m + 0R_{20} + m + 0R_{30} + m = 4m$$
$$m + 0R_{11} + m + R_{21}^{0} + m + 0R_{31} + m = 4m + 1$$
$$m + 0R_{12} + (m + 1) + 0R_{22} + (m + 1) + 0R_{32} + m = 4m + 2$$
$$m + R_{13}^{0} + m + R_{23}^{0} + m + R_{33}^{0} + m = 4m + 3$$

A rank R_{ir} appearing as $R_{ir}^{0} = 1$ in the segmentation identity implies that the rank is an integer, and a rank R_{ir} appearing as $0R_{ir} = 0$ implies that the corresponding rank is not an integer. It is seen that equisegmentation property is satisfied by the Popular Method for $r = 0,1,3$ but not for $r = 2$.

Method 2 (Popular Method with Arithmetic Rounding)

This method based on arithmetic rounding applied to the ranks is offered by the Popular Method. Ranks are given below for different sample sizes.

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

Segmentation identities are given by

$$(m-1)+R_{10}^{0}+m+R_{20}^{0}+(m-1)+R_{30}^{0}+(m-1)=4m$$
$$m+R_{11}^{0}+(m-1)+R_{21}^{0}+(m-1)+R_{31}^{0}+m=4m+1$$
$$m+R_{12}^{0}+m+R_{22}^{0}+(m-1)+R_{32}^{0}+m=4m+2$$
$$m+R_{13}^{0}+m+R_{23}^{0}+m+R_{33}^{0}+m=4m+3$$

It is seen that equisegmentation property is satisfied by the Popular Method only for $r=3$.

Method 3 (Mendenhall and Sincich Method) This method suggests to round up the rank of the first quartile provided by the popular method if the rank is halfway between two integers. It also suggests rounding down the rank of the third quartile if the rank is halfway between two integers. It is easy to see that the suggestion by Mendenhall and Sincich (1995, p54) only applies to samples with size $n=4m+1$. For other sample sizes ranks offered by the Popular Method do not lie exactly in the halfway between two integers, and as such those ranks are the same in both the Popular Method and the Mendenhall and Sincich Method. The ranks offered by the Mendenhall and Sincich Method are tabulated below.

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

Segmentation identities are given by

$$m+0R_{10}+m+0R_{20}+m+0R_{30}+m=4m$$
$$m+R_{11}^{0}+(m-1)+R_{21}^{0}+(m-1)+R_{31}^{0}+m=4m+1$$
$$m+0R_{12}+(m+1)+0R_{22}+(m+1)+0R_{32}+m=4m+2$$
$$m+R_{13}^{0}+m+R_{23}^{0}+m+R_{33}^{0}+m=4m+3$$

It is seen that equi-segmented property is satisfied by the Mendenhall and Sincich Method for $r=0,3$ but not for $r=1,2$.

Method 4 By this method, ranks of quartiles are given by $R_\alpha=\alpha\ n/4,\ \ \alpha=1,2,3$. Separate the largest integer (i) not exceeding R_α, and decimal part (d) of R_α and write $R_\alpha=i+d$. The quartile is finally given by

$$Q_\alpha=x_{(i)}+d\,(x_{(i+1)}-x_{(i)})=(1-d)\,x_{(i)}+d\,x_{(i+1)},\quad (\alpha=1,2,3)$$

where $x_{(i)}$ is the *i*-th observation. This method is a slight variation of Popular Method discussed in Section 2. Ranks are given below for different sample sizes.

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

Segmentation identities are given by

$$(m-1)+R_{10}^{0}+(m-1)+R_{20}^{0}+(m-1)+R_{30}^{0}+m=4m$$
$$m+0R_{11}+m+0R_{21}+m+0R_{31}+(m+1)=4m+1$$
$$m+0R_{12}+m+R_{22}^{0}+m+0R_{32}+(m+1)=4m+2$$
$$m+0R_{13}+(m+1)+0R_{23}+(m+1)+0R_{33}+(m+1)=4m+3$$

It is seen that equisegmentation property is not satisfied for any $r=0,1,2,3$.

Method 5 (Hines and Montgomery, 1990, 18) Ranks of quartiles are given by $R_\alpha = \alpha\ n/4+0.5,\ \ \alpha=1,2,3$. Separate the largest integer (i) not exceeding R_α, and decimal part (d) of R_α and write $R_\alpha = i+d$. The quartile is finally given by

$$Q_\alpha = x_{(i)} + d\,(x_{(i+1)} - x_{(i)}) = (1-d)\,x_{(i)} + d\,x_{(i+1)}, \quad (\alpha = 1,2,3)$$

where $x_{(i)}$ is the *i*-th observation. This method is a slight variation of Method 4. Ranks are given below for different sample sizes.

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

Segmentation identities are given by

$$m+0R_{10}+m+0R_{20}+m+0R_{30}+m=4m$$
$$m+0R_{11}+m+R_{21}^{0}+m+0R_{31}+m=4m+1$$
$$m+R_{12}^{0}+m+0R_{22}+m+R_{32}^{0}+m=4m+2$$
$$(m+1)+0R_{13}+m+R_{23}^{0}+m+0R_{33}+(m+1)=4m+3$$

It is seen that equisegmentation property is satisfied by this method for $r=0,1,2$ but not for $r=3$.

Method 6 (Johnson, 2000, p32) Ranks of quartiles are given by $R_\alpha = \alpha\ n/4,\ \ \alpha = 1, 2, 3$. Separate the largest integer (i) not exceeding R_α, and decimal part (d) of R_α and write $R_\alpha = i + d$. If $n/4$ is not an integer, round it up to the next integer and find the corresponding ordered observation. If $n/4$ is an integer, calculate the mean of the $(n/4)$ th and the next observation. Ranks are given below for different sample sizes.

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

Segmentation identities are given by

$$m + 0R_{10} + m + 0R_{20} + m + 0R_{30} + m = 4m$$
$$m + R_{11}^0 + (m-1) + R_{21}^0 + (m-1) + R_{31}^0 + m = 4m+1$$
$$m + R_{12}^0 + (m-1) + R_{22}^0 + m + R_{32}^0 + m = 4m+2$$
$$m + R_{13}^0 + m + R_{23}^0 + m + R_{33}^0 + m = 4m+3$$

It is seen that equisegmentation property is satisfied by this method for $r = 0,3$ but not for $r = 1,2$.

Method 7 (Hinge Method) The oldest and most popular method to find extreme quartiles is based on finding the median first, and then finding the medians of upper and lower halves of the data . The tradition is to count the median in both halves (Mayer and Sykes, 1996, 25). Tukey (1977, p32-35) called them hinges.

For $n = 4m + 0$, it follows that 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 $R_{30} = [(2m + 2/4) + 4m]/2 = 3m + 1/4$. For $n = 4m+1$, it follows that 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 $R_{31} = [(2m+1) + (4m+1)]/2 = 3m+1$.

For $n = 4m+2$, it follows that 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 $R_{32} = [(2m + 1 + 2/4) + (4m+2)]/2 = 3m + 1 + 3/4$. For $n = 4m+3$, it follows that 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 of the quartiles by the Hinge Method are summarized below:

	$n=4m$	$n=4m+1$	$n=4m+2$	$n=4m+3$
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}	$3m+1/4$	$3m+1$	$3m+1+3/4$	$3m+2+2/4$

Segmentation identities are given by

$$m+0R_{10}+m+0R_{20}+m+0R_{30}+m=4m$$
$$m+R_{11}^{0}+(m-1)+R_{21}^{0}+(m-1)+R_{31}^{0}+m=4m+1$$
$$(m+1)+0R_{12}+m+0R_{22}+m+0R_{32}+(m+1)=4m+2$$
$$(m+1)+0R_{13}+m+R_{23}^{0}+m+0R_{33}+(m+1)=4m+3$$

Clearly equisegmentation property is satisfied by the Hinge Method only for $r=0$.

Method 8 (Vinning Method) The formulae given by Vinning (1998, p 44) can be simplified as

$$Q_1=\begin{cases}(n+3)/4 \text{ th observation if } n \text{ is odd}\\(n+2)/4 \text{ th observation if } n \text{ is even}\end{cases}$$

$$Q_3=\begin{cases}(3n+1)/4 \text{ th observation if } n \text{ is odd}\\(3n+2)/4 \text{ th observation if } n \text{ is even}\end{cases}$$

The example he provides with $n=35$ divide the observations into four segments with 9, 8, 8 and 9 observations among them. The median has an integer rank namely the 18th position. Ranks are given below for different sample sizes.

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

Segmentation identities are given by

$$m + 0R_{10} + m + 0R_{20} + m + 0R_{30} + m = 4m$$
$$m + R_{11}^0 + (m-1) + R_{21}^0 + (m-1) + R_{31}^0 + m = 4m+1$$
$$m + R_{12}^0 + m + 0R_{22} + m + R_{32}^0 + m = 4m+2$$
$$(m+1) + 0R_{13} + m + R_{23}^0 + m + 0R_{33} + (m+1) = 4m+3$$

Clearly equisegmentation property is satisfied by the Vinning Method only for $r = 0$ and $r = 2$.

Method 9 (Smith Method) The formulae provided for percentiles by Smith (1997, pp36-38) can be specialized to quartiles as

$$Q_1 = \begin{cases} (n+2)/4 \text{ th observation if } n/4 \text{ is not an integer} \\ [(n/4 \text{ th} + \text{the next) observation }]/2 \text{ if } n/4 \text{ is an integer} \end{cases}$$

$$Q_3 = \begin{cases} (3n+2)/4 \text{ th observation if } 3n/4 \text{ is not an integer} \\ [(3n/4 \text{ th} + \text{the next) observation}]/2 \text{ if } 3n/4 \text{ is an integer} \end{cases}$$

He suggests rounding the ranks to the nearest integer. The example he provides for $n = 12$ does satisfy equisegmentation property with $m = 3$. Ranks are given below for different sample sizes.

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

Segmentation identities are given by

$$m + 0R_{10} + m + 0R_{20} + m + 0R_{30} + m = 4m$$
$$m + 0R_{11} + m + R_{21}^0 + m + 0R_{31} + m = 4m+1$$
$$m + R_{12}^0 + m + 0R_{22} + m + R_{32}^0 + m = 4m+2$$
$$(m+1) + 0R_{13} + m + R_{23}^0 + m + 0R_{33} + (m+1) = 4m+3$$

Clearly equisegmentation property is satisfied by the Vinning Method for $r = 0,1,2$ but not for $r = 3$.

Method 10 (Shao Method) It is surprising that the method proposed by (Shao, 1976, pp 174-175) is the only method in the literature that enjoys equisegmentation property.

(a) If the sample size is divisible by 4, the quartiles can be easily determined. When a quartile is located between two values, the mid point of them is considered to be the quartile.

(b) If the sample size is not divisible by 4, the quartiles can easily be determined in three steps:

(i) If the sample size is even, Q_1 is the median obtained from the lower 50% values of the sample.
(ii) If the sample size is odd, Q_1 is the median obtained from the lower 50% values of the sample after having discarded the median of the complete sample.

(iii) Locate Q_3 by the methods stated in *(i)* and *(ii)* except that the upper 50% of the values of the sample are used in the process.

Ranks are given below for different sample sizes.

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

Segmentation identities are given by

$$m+0R_{10}+m+0R_{20}+m+0R_{30}+m=4m$$
$$m+0R_{11}+m+R_{21}^{0}+m+0R_{31}+m=4m+1$$
$$m+R_{12}^{0}+m+0R_{22}+m+R_{32}^{0}+m=4m+2$$
$$m+R_{13}^{0}+m+R_{23}^{0}+m+R_{33}^{0}+m=4m+3$$

It is surprising that equisegmentation property is satisfied by the Shao Method for any value of r.

4. Suggested Methods

We discuss below two methods namely the Halving Method and the Remainder Method each of which satisfies equisegmentation property.

4.1 The Halving Method

We observe that Method 7 guarantees equisegmentation property if the median of the whole data set is always ignored in the calculation of the lower and upper quartiles. Method 1 with this kind of adjustment will hereafter be called the Halving Method (Joarder, 2002).

Example 4.1 We calculate below the quartiles of the data in Example 2.1 by the halving method. The rank of the median is $R_{22} = (1+n)/2 = 5.5$ so that $Q_{22} = x_{(5.5)} = (1-0.5)\, x_{(5)} + 0.5\, x_{(6)} = 0.5(3.9) + 0.5(4.7) = 4.3$. The first quartile is the median of the observations below the median of the whole sample i.e. $R_{12} = (1+5)/2 = 3$ so that $Q_{12} = x_{(3)} = 2$. The third quartile is the median of the observations above the median of the whole sample i.e. is $R_{32} = (6+10)/2 = 8$ so that $Q_{32} = x_{(8)} = 7.6$. To check the equisegmentation property, we show the position of the quartiles by downward arrows in the sample:

		$\Downarrow$			$\Downarrow$			$\Downarrow$		
1.7	1.9	2.0	2.8	3.9		4.7	6.2	7.6	12.1	29.3

We observe that there are $2(=m)$ observations in each of the four segments i.e. the quartiles do satisfy the equisegmentation property. The ranks of the quartiles given by the Halving Method are tabulated below:

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

4.2 The Remainder Method

We observe that each of the rank R_{10} and R_{30} given by the Halving Method is smaller than that given by the Remainder Method by $1/4$ .We also observe that the ranks of quartiles given by the Popular Method satisfy equisegmentation property if the rank is rounded down for $(r = 2, d = 1)$ and rounded up for $(r = 2, d = 3)$. Note that the remainder method for quartiles happens to be arithmetic rounding for outer quartiles when $r = 2$. A special kind of rounding applied to the ranks provided by the Popular Method for quantiles of even order has been discussed by Joarder (2002). The ranks thus obtained, called the Remainder Method, satisfy equisegmentation property.

Let $[u]$ be the largest integer not exceeding $u = u_{ir} = i(r+1)/4$, $\uparrow u \uparrow$ be the smallest integer exceeding u and $i(r+1) = d_{ir} \bmod 4$. Then we have the theorem.

Theorem 4.1 Let $m = (n-r)/4$, $n = 4m + r \geq 4$, and R_{ir} be the rank of the i th quartile with m observations in each segment. Then the ranks given by

$$R_{ir} = \begin{cases} im + [u_{ir}] \text{ if } (r,d) \in A, \text{and } d \le 2 & (4.1a) \\ im + \uparrow u_{ir} \uparrow \text{ if } (r,d) \in A, \text{and } d > 2 & (4.1b) \\ im + u_{ir} \quad \text{if } (r,d) \notin A & (4.1c) \end{cases}$$

where i and r are integers with $1 \le i \le 3$ and $0 \le r \le 3$, and $A = \{(r,d) : (2,1)(2,3)\}$ satisfy equisegmentation property. If $(r,d) \notin A$, the quartiles can be calculated by the simple linear interpolation as

$Q_{ir} = (1 - d/4)\, x_{(im+[u])} + (d/4)\, x_{(im+[u]+1)}$ where $x_{(i)}$ is the i th ordered observation.

Example 4.2 Let us now calculate quartiles for the sample in Example 2.1 by the remainder method. Here $n = 10 = 4(2) + 2$ i.e. $m = 2, r = 2$. Since $u_{12} = 1(2+1)/4 = 3/4$ (i.e. $r = 2, d = 3 > 2$), the rank of the first quartile is (see 4.1b)
$R_{12} = 1(m) + \uparrow u_{12} \uparrow = 2 + \uparrow 3/4 \uparrow = 3$. Again since $u_{22} = 2(2+1)/4 = 1 + 2/4$ (i.e. $r = 2, d = 2 \le 2$), the rank of the second quartile (see 4.1 c) is
$R_{22} = 2(m) + u_{22} = 2(2) + 1 + 2/4 = 5.5$. Finally since $u_{32} = 3(2+1)/4 = 2 + 1/4$ (i.e. $r = 2, d = 1 \le 2$), the rank of the third quartile (4.1 a) is $R_{32} = 3(m) + [u_{12}] = 6 + 2 + [1/4] = 8$.
So the quartiles are $Q_{12} = x_{(3)} = 2.0$, $R_{22} = (1-0.5)x_{(5)} + 0.5\, x_{(6)} = (3.9 + 4.7)/2 = 4.3$ and $R_{32} = x_{(8)} = 7.6$

To check the equisegmentation property, we show the position of the quartiles by downward arrows in the sample:

$$\begin{array}{cccccccccc} & & \Downarrow & & & \Downarrow & & \Downarrow & & \\ 1.7 & 1.9 & 2.0 & 2.8 & 3.9 & \; 4.7 & 6.2 & 7.6 & 12.1 & 29.3 \end{array}$$

We observe that the equisegmentation property is satisfied here with $m = 2$. The ranks of the quartiles given by the Remainder Method are tabulated below:

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

The Halving Method as well as the Remainder Method satisfies equisegmentation property. It is worth noting that in each of the two methods the value of r $(n = 4m + r)$ is the number of quartiles with integer ranks. The Shao Method, however, doesn't have algebraic expression for the ranks and may not be that suitable for using it or generalizing it to other quantiles. Though the Halving Method is the simplest one and satisfies equisegmentation property, it, seems, difficult to generalize the notion to quantiles in general. The Remainder Method is generalized to quartiles of even order by Joarder (2002a).

Acknowledgements

The authors acknowledge the excellent research facilities available at King Fahd University of Petroleum and Minerals, Saudi Arabia. The authors are grateful to Prof. M. M. Ali, Ball State University, USA for constructive suggestions that have improved the quality of the paper.

References

Bluman, A. G. (2001). *Elementary Statistics: A Step by Step Approach*. McGraw Hill, New York.

Hines, W. and Montgomery, D.C. (1990). *Probability and Statistics in Engineering and Management Sciences*. New York: John Wiley.

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

Joarder, A.H. (2002). The hinge method and the halving method for sample quartiles. *Technical Report No*. 273, King Fahd University of Petroleum and Minerals, Saudi Arabia.

Joarder, A.H. (2002a). The remainder method for sample quantiles of even order. *Technical Report No*. 274 , King Fahd University of Petroleum and Minerals, Saudi Arabia.

Johnson, R. (2000). *Miller and Freund's Probability and Statistics for Engineers*. Prentice Hall

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

Mayer, A.D. and Sykes, A.M. (1996). *Statistics*. London: Edward Arnold.

Mendenhall, W. and Sincich, T. (1995). *Statistics for Engineering and the Sciences*. Englewood Cliffs, NJ: Prentice Hall.

Ostle, B., Turner, K.V. Hicks, C.R. and McElrath, G.W. (1996). *Engineering Statistics*: *The Industrial Experience*. New York: Duxbury Press.

Shao, S.P. (1976). *Statistics for Business and Economics*. Charles E. Merill Publishing Co., Columbus, Ohio.

Smith, P.J. (1997). *Into Statistics*. Heidelberg: Springer.

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

Vinning, G.G. (1998). *Statistical Methods for Engineers*. New York: Duxbury Press.