A Comparison and Contrast of Some Methods for Sample Quartiles

Anwar H. Joarder and Raja M. Latif
King Fahd University of Petroleum & Minerals

ABSTRACT A remainder representation of the sample size $n = 4m + r$ ($r = 0, 1, 2, 3$) is exploited to write out the ranks of quartiles exhaustively which in turn help compare ranks for quartiles provided by different methods available in the literature. The criterion of the equisegmentation property that the number of integer ranks below the first quartile, that between the consecutive quartiles, and that above the third quartile are the same, has been used to compare and contrast different methods. Four segmentation identities can be 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. The quartiles provided by each of these two methods satisfy the equi-segmentation property if the observations are distinct. More interestingly, in these two methods r also represents the number of quartiles having integer ranks.

Keywords : 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 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 $P(X \le x_p) = p$ since $P(X = x_p) = 0$. 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

$$f(x) = \beta^{-1} e^{-x/\beta}, \qquad x > 0$$

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

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

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

$$\begin{aligned} P(X < 1) &= P(X = 0) = 0.0625 \le 0.25, \\ P(X \le 1) &= P(X = 0) + P(X = 1) = 0.3125 \ge 0.25. \end{aligned}$$

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

$$\begin{aligned} P(X < 2) &= 0.3125 \le 0.50, \\ P(X \le 2) &= 0.6825 \ge 0.50 \end{aligned} .$$

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 $P(X < 2) = 0.3125 \le 0.60$ and $P(X \le 2)$ $= 0.6825 \ge 0.60$. 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 [6], p. 52).

A remainder representation of the sample size $n = 4m + r$ $(r = 0, 1, 2, 3)$ is exploited to write out the ranks of quartiles exhaustively which in turn help compare ranks for quartiles provided by different methods available in the literature. Some of them differ only by 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 integer ranks below 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 clearly show 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. The 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 the Popular Method hereinafter, is described below. The rank of the i(i = 1, 2, 3)th quartile is given by

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

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

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

where $x_{(l)}$ is the lth ordered observation (Ostle *et al*. [12], p. 38). To write out the ranks exhaustively let us denote the sample size $n(\geq 4)$ by the following remainder-modulus representation

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

so that the number of observations in each of the four segments is given by $m = (n-r)/4$. With this representation of the sample size, the ranks and the 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 equi-segmentation property guarantees that $m_1 = m_2 = m_3 = m_4$ if the observations are distinct. In case, $1 \leq n \leq 3$, 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 $n = 4m + r\ \ (m \geq 1;\ r = 0,1,3)$ if the observations are distinct. For $n = 4m + 2\ \ (m \geq 1)$, the number of observations in four segments are given by

$m, (m+1), (m+1)$ and m respectively if the observations are distinct.

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) are given below:

1.7 1.9 2.0 2.8 3.9 4.7 6.2 7.6 12.1 29.3

(Bluman [1], p.137). We now calculate quartiles by the Popular Method. Here the sample size is $n = 10 = 4(2) + 2$ so that $m = 2$ and $r = 2$. For $r = 2$ we will denote the ranks of quartiles by R_{i2} $(i = 1, 2, 3)$. The ranks of the quartiles provided by the Popular Method are (see equation 2.1) given by

$$R_{12} = (n+1)/4 = 2.75, \ R_{22} = (n+1)/2 = 5.5, \ R_{32} = 3(n+1)/4 = 8.25$$

so that by linear interpolation (see equation 2.2) the quartiles are given by

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

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

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

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

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

where $x_{(i)}$ is the ith ordered observation, $u = u(i,r) = i(r+1)/4$, $[u]$ is the greatest integer not exceeding 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):

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

The ranks for different sample sizes provided by this 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}	$m + 1/4$	$m + 2/4$	$m + 3/4$	$m + 1$
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 the equisegmentation property is satisfied by the Popular Method for r = 0, 1, 3 but not for $r = 2$. Note that the Lapin Method (Lapin [7], 45-46) is a representation of the Popular Method accommodating simple linear interpolation.

Method 2 (**Popular Method with Arithmetic Rounding**) This method is based on arithmetic rounding applied to the ranks offered by the Popular Method. The ranks for different sample sizes provided by this method are tabulated below:

R_{ir}	$n = 4m$ $r = 0$	$n = 4m+1$ $r = 1$	$n = 4m+2$ $r = 2$	$n = 4m+3$ $r = 3$
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 the 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 ([9], p. 54) 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 for different sample sizes provided by this method are tabulated below:

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

Segmentation identities are given by

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

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

Method 4 By this method, the ranks of quartiles are given by $R_\alpha = \alpha n/4, \quad \alpha = 1, 2, 3$. Separate the largest integer (i) not exceeding R_α, and decimal part (d) of R_α and write $R_\alpha = i + d$. The α th $(\alpha = 1, 2, 3)$ 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 ith observation. This method is a slight variation of the Popular Method discussed in Section 2. The ranks for different sample sizes provided by this 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	$m+1/4$	$m+1/2$	$m+3/4$
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

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

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

Method 5 (Hines and Montgomery [2], p. 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 α th $(\alpha = 1, 2, 3)$ 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 ith observation. This method is a slight variation of Method 4. The ranks for different sample sizes provided by this 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}	$m + 1/2$	$m + 3/4$	$m + 1$	$m + 1 + 1/4$
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 the equisegmentation property is satisfied by this method for r = 0, 1, 2 but not for r = 3.

Method 6 (Johnson [5], p. 32) The ranks of the 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 corres- ponding ordered observation. If $n/4$ is an integer, calculate the mean of the $(n/4)$th and the next observation. The ranks for different sample sizes provided by this 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}	$m + 1/2$	$m + 1$	$m + 1$	$m + 1$
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 the equisegmentation property is satisfied by this method for r = 0, 3 but not for r = 1, 2.

Method 7 (Hinge Method) An interesting 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 [8], p. 25). Tukey ([16], pp. 32-35) called them hinges.

For $n = 4m$, 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 for different sample sizes provided by this 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}	$m+3/4$	$m+1$	$m+1+1/4$	$m+1+2/4$
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 the equisegmentation property is satisfied by the Hinge Method only for $r = 0$.

Method 8 (Vinning Method) The formulae given by Vinning ([17], 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 $n = 35$ divides the ordered sample observations into four segments with 9, 8, 8 and 9 observations among them. The median has an integer rank namely the 18th position. The ranks for different sample sizes provided by this method are tabulated below:

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

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 the equisegmentation property is satisfied by the Vinning Method only for $r=0$ and $r=2$. Milton and Arnold Method ([11], pp. 207-208) suggested the ranks of extreme quartiles to be $R_{1r}=([(n+1)/2]+1)/2$ and $R_{3r}=n+1-R_1$ but it turns out that they are exactly the same as the ranks of extreme quartiles given by the Vinning Method.

Method 9 (**Siegel Method**) Siegel ([14], p. 117) suggests the ranks of extreme quartiles to be $R_{1r}=([(n+1)/2]+1)/2$ and $R_{3r}=n+1-R_1$ while, unlike any other method, he suggests the rank of the median to be $R_{2r}=[(n+1)/2]$ where $[a]$ is the largest integer not exceeding a. The ranks for different sample sizes provided by this method are tabulated below:

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

Segmentation identities are given by

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

Clearly the equisegmentation property is not satisfied for any value of r .

Method 10 (**Smith Method**) The formulae provided for percentiles by Smith ([15], pp. 36-38) can be specialized to quartiles as

$$Q_1 = \begin{cases} \dfrac{n+2}{4} \text{ th observation if } n/4 \text{ is not an integer} \\ \dfrac{1}{2}\left(\dfrac{n}{4}\text{th } + \dfrac{n+4}{4}\text{th }\right) \text{observation} \quad \text{if } n/4 \text{ is an integer} \end{cases}$$

$$Q_3 = \begin{cases} \dfrac{3n+2}{4} \text{ th observation if } 3n/4 \text{ is not an integer} \\ \dfrac{1}{2}\left(\dfrac{3n}{4}\text{th } + \dfrac{3n+4}{4}\text{th}\right) \text{observation} \quad \text{if } n/4 \text{ is an integer} \end{cases}$$

He suggests rounding the ranks to the nearest integer. The example he provides for $n = 12$ does satisfy the equisegmentation property with $m = 3$. The ranks for different sample sizes provided by this 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}	$m+1/2$	$m+3/4$	$m+1$	$m+1+1/4$
R_{2r}	$2m+2/4$	$2m+1$	$2m+1+2/4$	$2m+2$
R_{3r}	$3m+2/4$	$3m+1+1/4$	$3m+2$	$3m+2+3/4$

Segmentation identities are given by

$$\begin{aligned} &m + 0R_{10} + m + 0R_{20} + m + 0R_{30} + m = 4m, \\ &m + 0R_{11} + m + R_{21}^{0} + m + 0R_{31} + m = 4m+1, \\ &m + R_{12}^{0} + m + 0R_{22} + m + R_{32}^{0} + m = 4m+2, \\ &(m+1) + 0R_{13} + m + R_{23}^{0} + m + 0R_{33} + (m+1) = 4m+3. \end{aligned}$$

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

Method 11 (**Shao Method**) It is surprising that the method proposed by Shao ([13], 1976, pp.174-175) is the only method in the literature that enjoys the 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 these two values 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:
   (1) If the sample size is even, Q_1 is the median obtained from the lower 50% values of the sample.

(2) 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.

(3) Locate Q_3 by the methods stated in (1) and (2) except that the upper 50% of the values of the sample are used in the process.

The ranks for different sample sizes provided by this 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}	$m+2/4$	$m+2/4$	$m+1$	$m+1$
R_{2r}	$2m+2/4$	$2m+1$	$2m+1+2/4$	$2m+2$
R_{3r}	$3m+2/4$	$3m+1+2/4$	$3m+2$	$3m+3$

Segmentation identities are given by

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

Observe that the equisegmentation property is satisfied by this 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 the equisegmentation property.

4.1 The Halving Method

We observe that Method 7 guarantees the 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 hereinafter be called the Halving Method (Joarder [3]).

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. $R_{32}=(6+10)/2=8$ 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 for different sample sizes provided by this method are tabulated below:

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

It may be remarked here that the first quartile is the median of the smallest $n/2$ observations if n is even, and that of the smallest $(n-1)/2$ observations if n is odd. Similarly the third quartile is the median of the largest $n/2$ observations if n is even, and that of the largest $(n-1)/2$ observations if n is odd.

4.2 The Remainder Method

We observe that each of the ranks 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 the quartiles given by the Popular Method satisfy the equisegmentation property if the rank is rounded down for $(r=2, d=1)$ and rounded up for $(r=2, d=3)$. A special kind of rounding applied to the ranks provided by the Popular Method for quantiles of even order has been discussed by Joarder [4]. The ranks obtained this way, called the Remainder Method, satisfy the equisegmentation property.

Let $[u]$ be the largest integer not exceeding u, and $\lceil u \rceil$ the smallest integer exceeding u. Again let $u = u_{ir} = i(r+1)/4$ and $i(r+1) = d_{ir} (\mathrm{mod}\, 4)$. Then we have the following theorem.

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

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

where i and r are integers with $1 \le i \le 3$ and $0 \le r \le 3$, and $A = \{(r,d) : r = 2,\ d = 1,3\}$, satisfy the equisegmentation property. If $(r,d) \notin A$, then 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 ith ordered observation.

Example 4.2 Let us now calculate the 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 $R_{12} = 1(m) + \lceil u_{12} \rceil = 2 + \lceil 3/4 \rceil = 3$ (see 4.1b). 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 is $R_{22} = 2(m) + u_{22} = 2(2) + 1 + 2/4 = 5.5$ (see 4.1 c). 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 is $R_{32} = 3(m) + [u_{12}] = 6 + 2 + [1/4] = 8$ (See 4.1a). 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:

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

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

R_{ir}	$n = 4m$ $r = 0$	$n = 4m+1$ $r = 1$	$n = 4m+2$ $r = 2$	$n = 4m+3$ $r = 3$
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 the 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 hence may not be suitable for using it or generalizing it to other quantiles. Though the Halving Method is the simplest one and satisfies the equisegmentation property, it seems to be difficult to generalize the notion to quantiles in general. Note that the Remainder Method for quartiles happens to be the Popular Method with arithmetic rounding for outer quartiles when $r = 2$. The Remainder Method is generalized to quantiles of even order by Joarder [4].

It remains open to check the equisegmentation property for samples with ties. Finally it is worth remarking that for a sample of large size, the empirical cumulative distribution function may be used to calculate sample quartiles (Mendenhall *et al*. [10], Section 15.1.1) or Ross, S. (1987, Section 4).

Acknowledgements

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

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

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

[3] Joarder, A. H. (2003). The halving method for sample quartiles, *International Journal of Mathematical Education in Science and Technolog*, 34(4), 629-633.

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

[5] Johnson, R. (2000). Miller and Freund’s Probability and Statistics for Engineers. Prentice Hall.

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

[7] Lapin, L. L. (1997). Modern Engineering Statistics, Duxbury Press.

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

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

[10] Mendenhall, W., Beaver, R. J. and Beaver, B. M. (2002). A Brief Introduction to Probability and Statistics, United States: Duxbury.

[11] Milton, J. S. and Arnold, J. C. (2003). Introduction to Probability and Statistics, New York: McGraw-Hill.

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

[13] Shao, S. P. (1976). Statistics for Business and Economics, Columbus, Ohio: Charles

E. Merill Publishing Co.
[14] Siegel, A. F. (2003). Practical Business Statistics, New York :McGraw-Hill.
[15] Smith, P. J. (1997). Into Statistics, Singapore: Springer-Verlag.
[16] Tukey, J. W. (1977). Exploratory Data Analysis, Reading, MA: Addison Wesley.
[17] Vinning, G. G. (1998). Statistical Methods for Engineers, New York: Duxbury Press.

Ross, S. (1987). Introduction to Probability and Statistics for Engineers and Scientists. New York: John Wiley.

A Comparison and Contrast of Some Methods for Sample Quartiles

Anwar H. Joarder and Raja M. Latif
King Fahd University of Petroleum & Minerals

There are about a dozen methods to find sample quartiles. The emergence of so many methods is due to non-rigorous definition of quartiles. In this talk we probe the issue, and suggest a new criterion of equisegmentation to determine quartiles. The existing methods have then been checked in the light of this criterion and found that only Shao Method satisfies it. Two new methods have been proposed and illustrated.