King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 343

December 2006

The Remainder Method for Sample Percentiles

A.H. JOARDER and M.R. ABUJIYA

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

The Remainder Method for Sample Percentiles

A.H. JOARDER and M.R. ABUJIYA

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

Abstract A method called the Remainder Method is proposed for the calculation of sample quantiles of a given order, for example, quartiles, hexatiles, octatiles, deciles and percentiles assuming that all the observations are distinct. Proof is given for a special case of deciles. We propose the criterion of equisegmentation property that the number of observations below the first quantile, that between the consecutive quantiles, and that above the last quantile are the same. The formulae for quantiles offered by the proposed method satisfy the equisegmentation property, and more interestingly provide the number of quantiles having integer ranks. Some open problems are indicated.

AMS Mathematics Subject Classification: *62-01*, *62-02*

Key Word and Phrases: Sample quartiles, deciles, percentiles, remainder method

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

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

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
$P(X<1)=P(X=0)=0.0625\le 0.25,\ P(X\le 1)=P(X=0)+P(X=1)=0.3125\ge 0.25.$
Similarly $x_{0.50}=2$, is the second quartile of the distribution since
$P(X<2)=0.3125\le 0.50,\quad P(X\le 2)=0.6825\ge 0.50$.

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 $P(X<2)=0.3125\le 0.60$ 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. 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 hinges (Tukey, 1976), is based on finding the median first and then finding the medians of the upper and lower halves of the data each time including the median of the whole data set. Done so, approximately 25% observations remain below the lower quartile and 25% above the upper quartile. The literature is full of different formulae for sample quartiles with various rounding notions of the corresponding ranks of quartiles. See for example Mendenhall and Sincich (1995, 54), and Joarder and Latif (2004) for a detailed survey and illustrations. Joarder (2003) discussed halving method of sample quartiles that satisfies equisegmentation property but it seems rather difficult to generalize it to quantiles of higher order.

In this note, the Remainder Method discussed by Joarder and Latif (2004) for quartiles has been generalized to some even orders namely hexatiles, octatiles, deciles and percentiles. Proof is given for a special case of deciles.

We propose the criterion of equisegmentation property that the number of observations below the first quantile, that between the consecutive quantiles, and that above the last quantile are the same. Let the number of observations in each segment be $m_i\quad(i=1,2,3,...,f\,;\,f=2,4,6,\cdots)$. Then the equisegmentation property requires that $m_1=m_2=\cdots=m_f$. However this will divide the ordered sample observations into the desired number of segments leaving the same number of observations in each if the observations are distinct.

Consider the quantiles of even order say $f = 2, 4, 6, ...,$ that divides the ordered sample observations in f divisions with $m(\geq 1)$ observations in each segment. Since the sample size can be represented by $n = r \pmod{f}$ i.e.

$$n = fm + r, \; (r = 0, 1, 2, \cdots, f - 1), \qquad (1.1)$$

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

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

or m for short.

The equisegmentation Property: The ranks R_{ir} $(i = 1, 2, \cdots, f - 1;\; r = 0, 1, \cdots, f - 1)$ for quartiles of order f satisfies equisegmentation property if

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

$$(ii)\; \lceil R_{ir} \rceil - \lfloor R_{i-1,r} \rfloor - 1 = m, \;\; i = 2, 3, \cdots, f \qquad (1.3b)$$

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

where $\lfloor x \rfloor$ and $\lceil x \rceil$ are the floor function (greatest 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 above the third quantile is m. The equation $(1.3b)$ states that the number of observations between two consecutive quantiles is m. Interestingly the quantity r is also the number of quantiles with integer ranks.

The Remainder Method for quartiles $(f = 4)$, hexatiles $(f = 6)$, octatiles $(f = 8)$ and deciles $(f = 10)$ have been discussed in Section 2. The method has been proved for a special case of deciles in Section 2.4, and argued that proofs for all other cases are similar. A set of general formulae for quantiles of some even order, in particular, or quantiles of any order, in general, remains open to be determined. The left hand side of the equation in (1.3b) in general, for any two numbers u and $v(<u)$, can be written as $\lceil v \rceil - \lfloor u \rfloor - 1$, which is actually the number of integers between u and $v(<u)$.

Sample quartiles are popularly interpolated linearly by the observations corresponding to the ranks $i(n+1)/4$, $(i = 1,2,3)$. This method will hereinafter be called the Popular Method. Joarder (2003) observed that the ranks provided by this method do not satisfy equisegmentation property if sample sizes are $n = 6,10,14$ etc. This led us to conjecture that the remainder of the sample size with respect modulus 4 may play a role in the determination of the ranks for quartiles.

Let R_{ir} be the rank of i^{th} quartile with m observations in each of the f segments. Then

$$R_{ir} = i(n+1)/f \quad = (fm + r + 1)/f \;= im + \; i(r+1)/f \;= im + \lfloor u_{ir} \rfloor + d/f \qquad (1.4)$$

where i and r are integers with $1 \le i \le f-1$, $0 \le r \le f-1$, $\lfloor u_{ir} \rfloor$ is the greatest integer not exceeding (less than or equal to) $u_{ir} = i(r+1)/f = \lfloor u_{ir} \rfloor + d/f$. For simplicity we will often use u for u_{ir}. The quartiles can then be calculated by the simple linear interpolation as

$$Q_{ir} = (1 - d/f)\, x_{(im+\lfloor u \rfloor)} + (d/f)\, x_{(im+\lfloor u \rfloor+1)}, \quad (i = 1, 2, \cdots, f-1; r = 0, 1, 2, \cdots, f-1) \qquad (1.5)$$

where $x_{(i)}$ is the i th ordered observation. Note that $u = d/f$ if $\lfloor u_{ir} \rfloor = 0$ i.e. if $u < 1$.

2. Sample quartiles, hexatiles, octatiles and deciles

2.1 The Remainder Method for Sample Quartiles $(f = 4)$

The refinement of the formulae for quartiles is based on the equisegmentation property discussed in Section 1. With a view to improving upon the rank of quartiles given by the Popular Method so that equisegmentation property is satisfied, a special notion of rounding depending on the remainder r and d of the ranks of quartiles considered by Joarder and Latif (2004) is given below.

Theorem 2.1 Let R_{ir} $(1 \le i \le 3;\ 0 \le r \le 3)$ be the rank of the i th quartile based on a sample of $n = 4m + r \ge 4$ observations. Then the following ranks satisfy equisegmentation property:

$$R_{ir} = \begin{cases} im + \lfloor u_{ir} \rfloor \ \text{ if } (r,d) \in A, \text{ and } d \le 2 & (2.1a) \\ im + \lceil u_{ir} \rceil \ \text{ if } (r,d) \in A, \text{ and } d > 2 & (2.1b) \\ im + u_{ir} \quad \text{ if } (r,d) \notin A & (2.1c) \end{cases}$$

where i and r are integers with $1 \le i \le 3$ and $0 \le r \le 3$, $u_{ir} = i(r+1)/4 = \lfloor u_{ir} \rfloor + d/4$,

and $A = \{(r,d) : (2,\ 1),\ (2,\ 3)\}$ with $m = (n-r)/4 (\ge 1)$ observations in each segment.

In what follows we will see that the ranks for quartiles satisfy the equisegmentation property for an admissible set $A = \{(r,d)\}$.

Example 2.1 An independent consumer group tested radial tires from a major brand to determine expected tread life. The data (in thousands of miles) are given below:

50	54	52	47	61	54.5	50.5	51
48	55	53	43	56	58	42	

(cf. Vinning, 1998, 193). The ordered sample observations are given by

42	43	47	48	50	50.5	51	52
53	54	54.5	55	56	58	61	

To illustrate the proposed method we make new data sets with the first $n = 12$, $n = 13$, $n = 14$, $n = 15$ observations labeling them as Data 1, Data 2, Data 3 and Data 4 respectively. It can be checked that the popular method satisfies equisegmentation property for all the above data sets except Data 3. We show below how the Remainder Method can be applied for Data 3 for sample quartiles so that it satisfies the equisegmentation property.

Here the sample size is $n = 14 = 4(3) + 2$ i.e. $m = 3, r = 2$. By Theorem 2.1, the ranks for quartiles are given by R_{12}, R_{22}, R_{32}. Since $u_{12} = 1(2+1)/4 = 3/4$, $(r,d) = (2,3) \in A$ and $d = 3 > 2$, it follows from $(2.1b)$ that $R_{12} = 1(m) + \lceil u_{12} \rceil = 3 + 1$. Again since $u_{22} = 2(2+1)/4 = 1 + 2/4$ and $(r,d) = (2,2) \notin A$, it follows from $(2.1c)$ that $R_{22} = 2m + u_{22} = 6 + 1 + 2/4 = 7 + 2/4$. Also since $u_{32} = 3(2+1)/4 = 2 + 1/4$, $(r,d) = (2,1) \in A$ and $d = 1 < 2$, it follows from $(2.1a)$ that $R_{32} = 3m + \lfloor u_{32} \rfloor = 9 + 2$. Thus by the Remainder Method ranks for quartiles are given by $R_{12} = 4,\ R_{22} = 7 + 2/4,\ R_{32} = 11$. Clearly the ranks satisfy equisegmentation property. The positions of quartiles for Data 3 given by

$Q_{12} = 4$ th obs $= 48$.

$Q_{22} = (7 + 2/4)$ th obs $= (1 - 2/4)$ (7 th obs) $+ (2/4)$ (8 th obs) $= 0.50\ (51) + 0.50\ (52) = 51.5$

$Q_{32} = 11$ th obs $= 54.5$

are : $x_{(1)}, x_{(2)}, x_{(3)}, \overset{\Downarrow}{x_{(4)}}, x_{(5)}, x_{(6)}, x_{(7)}, \Downarrow x_{(8)}, x_{(9)}, x_{(10)}, \overset{\Downarrow}{x_{(11)}}, x_{(12)}, x_{(13)}, x_{(14)}$

where $x_{(i)}$'s are the ordered sample observations. The quartiles Q_{ir} $(i = 1,2,3)$ for a particular r are the quartiles popularly denoted by Q_1, Q_2, Q_3. The above rounding of ranks guarantees the desirable equisegmentation property. There are $m = 3$ observations in each segment here. The remainder $r (= 2$ here$)$ is also, as expected, the number of quartiles having integer ranks for any sample of size $n \geq 4$.

2.2 The Remainder Method for Hexatiles $(f = 6)$

Hexatiles are five numbers that divide ordered sample observations into six segments. The following theorem guarantees that the ranks for hexatiles given by the Remainder Method satisfy the equisegmentation property. The proof of the theorem is omitted as a more general and popular case of deciles $(f = 10)$ is proved in Theorem 2.4.

Theorem 2.2 Let R_{ir} $(1 \leq i \leq 5\ ;\ 0 \leq r \leq 5)$ be the rank of the ith hexatile based on n $n = 6m + r \geq 6$ observations. Then the following ranks of hexatiles will satisfy the equisegmentation property:

$$R_{ir} = \begin{cases} im + \lfloor u_{ir} \rfloor \text{ if } (r,d) \in A, \text{ and } d \leq 3 & (2.2a) \\ im + \lceil u_{ir} \rceil \text{ if } (r,d) \in A, \text{ and } d > 3 & (2.2b) \\ im + u_{ir} \text{ if } (r,d) \notin A & (2.2c) \end{cases}$$

where i and r are integers with $1 \leq i \leq 5$ and $0 \leq r \leq 5$, $u_{ir} = i(r+1)/6 = \lfloor u_{ir} \rfloor + d/6$ and an admissible set $A = \{(r,d) : (3,2), (4,1), (4,2), (4,4), (4,5)\}$ with $m = (n - r)/6 \geq 1$ observations in each segment.

Example 2.2 Consider Data 4 with sample size $n = 15 = 6(2) + 3$ i.e. $m = 2,\ r = 3$. The ranks for the hexatiles are given by $R_{13}, R_{23}, R_{33}, R_{43}, R_{53}$. Since $u_{13} = 1(3+1)/6 = 4/6$,

$(r,d)=(3,4)\notin A$, it follows from (2.2c) that $R_{13}=1(m)+u_{13}=2+4/6$. Again, since $u_{23}=2(3+1)/6=1+2/6$ and $(r,d)=(3,2)\in A$ and $d<3$, it follows from $(2.2a)$ that $R_{23}=2m+\lfloor u_{23}\rfloor=2(2)+1=5$. Also since $u_{33}=3(3+1)/6=2+0/6$, $(r,d)=(3,0)\notin A$, it follows from $(2.2c)$ that $R_{33}=3m+u_{33}=3(2)+2=8$. Thus by the Remainder Method ranks for hexatiles are given by $R_{13}=2+4/6,\ R_{23}=5,\ R_{33}=8, R_{43}=10+4/6, R_{53}=13$. Clearly the ranks satisfy equisegmentation property. The positions of hexatiles for Data 4 given by
$Q_{13}=(2/6)(43)+(4/6)(47)\approx 45.67,\ Q_{23}=50,\ \text{Q}_{33}=52$,
$Q_{43}=(2/6)(54)+(4/6)(54.5)\approx 54.33,\ Q_{53}=56$

are: $x_{(1)}, x_{(2)}, \Downarrow x_{(3)}, x_{(4)}, \overset{\Downarrow}{x}_{(5)}, x_{(6)}, x_{(7)}, \overset{\Downarrow}{x}_{(8)}, x_{(9)}, x_{(10)}, \Downarrow x_{(11)}, x_{(12)}, \overset{\Downarrow}{x}_{(13)}, x_{(14)}, x_{(15)}$

where $x_{(i)}$'s are the ordered sample observations. The above rounding of ranks guarantees the desirable equisegmentation property. There are $m=2$ observations in each segment here. The remainder r (which is 3 here) is also, as expected, the number of hexatiles having integer ranks for any sample of size $n\geq 6$.

2.3 The Remainder Method for Octatiles $(f=8)$

Octatiles are seven numbers that divide ordered sample observations into eight segments. The following theorem guarantees that the ranks for octatiles given by the Remainder Method satisfy equisegmentation property.

Theorem 2.3 Let R_{ir} $(1\leq i\leq 7;\ 0\leq r\leq 7)$ be the rank of the ith octatile based on $n=8m+r\geq 8$ observations. Then the following ranks for octatiles satisfy the equisegmentation property:

$$R_{ir}=\begin{cases} im+\lfloor u_{ir}\rfloor & \text{if } (r,d)\in A, \text{ and } d\leq 4 & (2.3a)\\ im+\lceil u_{ir}\rceil & \text{if } (r,d)\in A, \text{ and } d>4 & (2.3b)\\ im+u_{ir} & \text{if } (r,d)\notin A & (2.3c)\end{cases}$$

where i and r are integers with $1\leq i\leq 7$ and $0\leq r\leq 7$, $u_{ir}=i(r+1)/8=\lfloor u_{ir}\rfloor+d/8$ and an admissible set
$A=\{(r,d):(2,1),(2,2),(4,1),(4,2),(4,3),(4,4),(5,2),(5,6),(6,1),(6,2),(6,3),(6,5),(6,6),(6,7)\}$ with $m=(n-r)/8\geq 1$ observations in each segment.

2.4 The Remainder Method for Deciles $(f=10)$

Deciles are nine numbers that divide ordered sample observations into ten segments. The following theorem guarantees that the ranks for deciles given by the Remainder Method satisfy the equisegmentation property.

Theorem 2.4 Let R_{ir} $(1\leq i\leq 9;0\leq r\leq 9)$ be the rank of the ith decile based on $n=10m+r\geq 10$ observations. Then the following ranks for deciles satisfy equisegmentation property:

$$R_{ir} = \begin{cases} im + \lfloor u_{ir} \rfloor \text{ if } (r,d) \in A, \text{ and } d \le 5 & (2.4a) \\ im + \lceil u_{ir} \rceil \text{ if } (r,d) \in A, \text{ and } d > 5 & (2.4b) \\ im + u_{ir} \text{ if } (r,d) \notin A & (2.4c) \end{cases}$$

where i and r are integers with $1 \le i \le 9$, $0 \le r \le 9$, $u_{ir} = i(r+1)/10 = \lfloor u_{ir} \rfloor + d/10$ and an admissible set

$$A = \{(r,d): (2,1), (2,2), (3,2), (5,2), (5,4), (6,1), (6,2), (6,3), (6,4)(6,8), (6,9), (7,2), (7,4), (7,8), (8,1), (8,2), (8,3), (8,4), (8,6), (8,7), (8,8), (8,9)\}$$

with $m = (n-r)/10 \ge 1$ observations in each segment.

Proof. By writing out the ranks for deciles by (1.1) with $f = 10$, it is easy to observe that no rounding is needed for $r = 0,1,4,9$ i.e. the ranks are given by 2.4 (c). For other cases of $r = 2,5,6,7,8$, some ranks need to be rounded so that the deciles satisfy equisegmentation property. Since proofs are similar in all cases of $r = 2,5,6,7,8$, we prove the theorem for a special case say $r = 6$. Let $n = 10(m) + 6$ so that $r = 6$. Then by Theorem 2.4, the ranks for deciles are given by

$R_{16} = 1(m) + 1(6+1)/10 = m + 7/10,$ since $d = 7, (r,d) = (6,7) \notin A$

$R_{26} = 2m + \lfloor 2(6+1)/10 \rfloor = 2m + \lfloor 1 + 4/10 \rfloor = 2m + 1,$ since $d = 4 < 5,$ $(r,d) = (6,4) \in A$

$R_{36} = 3m + \lfloor 3(6+1)/10 \rfloor = 3m + \lfloor 2 + 1/10 \rfloor = 3m + 2,$ since $d = 1 < 5,$ $(r,d) = (6,1) \in A$

$R_{46} = 4m + \lceil 4(6+1)/10 \rceil = 4m + \lceil 2 + 8/10 \rceil = 4m + 3,$ since $d = 8 > 5,$ $(r,d) = (6,8) \in A$

$R_{56} = 5m + 5(6+1)/10 = 5m + 3 + 5/10,$ since $d = 5 \le 5,$ $(r,d) = (6,5) \notin A$

$R_{66} = 6m + \lfloor 6(6+1)/10 \rfloor = 6m + 4,$ since $d = 2 < 5,$ $(r,d) = (6,2) \in A$

$R_{76} = 7m + \lceil 7(6+1)/10 \rceil = 7m + 5,$ since $d = 9 > 5, (r,d) = (6,9) \in A$

$R_{86} = 8m + 8(6+1)/10 = 8m + 5 + 6/10,$ since $d = 6, (r,d) = (6,6) \notin A$

$R_{96} = 9m + \lfloor 9(6+1)/10 \rfloor = 9m + 6,$ since $d = 3 < 6, (r,d) = (6,3) \in A$.

Then it is easy to check from the above that

(i) $\lceil R_{16} \rceil - 1 = \lceil m + 7/10 \rceil - 1 = (m+1) - 1 = m$

$$
\begin{aligned}
(ii)\quad & \lceil R_{26}\rceil - \lfloor R_{16}\rfloor - 1 = \lceil 2m+1\rceil - \lfloor m+7/10\rfloor - 1 = (2m+1)-(m)-1 = m,\\
& \lceil R_{36}\rceil - \lfloor R_{26}\rfloor - 1 = \lceil 3m+2\rceil - \lfloor 2m+1\rfloor - 1 = 3m+2-(2m+1)-1 = m,\\
& \lceil R_{46}\rceil - \lfloor R_{36}\rfloor - 1 = \lceil 4m+3\rceil - \lfloor 3m+2\rfloor - 1 = 4m+3-(3m+2)-1 = m,\\
& \lceil R_{56}\rceil - \lfloor R_{46}\rfloor - 1 = \lceil 5m+3+5/10\rceil - \lfloor 4m+3\rfloor - 1 = 5m+4-(4m+3)-1 = m,\\
& \lceil R_{66}\rceil - \lfloor R_{56}\rfloor - 1 = \lceil 6m+4\rceil - \lfloor 5m+3+5/10\rfloor - 1 = 6m+4-(5m+3)-1 = m,\\
& \lceil R_{76}\rceil - \lfloor R_{66}\rfloor - 1 = \lceil 7m+5\rceil - \lfloor 6m+4\rfloor - 1 = 7m+5-(6m+4)-1 = m,\\
& \lceil R_{86}\rceil - \lfloor R_{76}\rfloor - 1 = \lceil 8m+5+6/10\rceil - \lfloor 7m+5\rfloor - 1 = 8m+6-(7m+5)-1 = m,\\
& \lceil R_{96}\rceil - \lfloor R_{86}\rfloor - 1 = \lceil 9m+6\rceil - \lfloor 8m+5+6/10\rfloor - 1 = 9m+6-(8m+5)-1 = m\\
(iii)\quad & 10m+6-\lfloor R_{96}\rfloor = 10m+6-\lfloor 9m+6\rfloor = 10m+6-(9m+6) = m\ .
\end{aligned}
$$

Thus it is proved that ranks of deciles given by the Remainder Method satisfy the equisegmentation property.

The remainder r (which is 6 here) is also, as expected, the number of deciles having integer ranks for any sample of size $n \geq 10$.

3. The Remainder Method for Percentiles $(f = 100)$

Percentiles are ninety-nine numbers that divide ordered sample observations into one hundred segments. The following theorem guarantees that the ranks for percentiles given by the Remainder Method satisfy the equisegmentation property.

Theorem 3.1 Let R_{ir} $(1 \leq i \leq 99; 0 \leq r \leq 99)$ be the rank of the i^{th} percentile based on $n = 100m + r \geq 100$ observations. Then the following ranks for percentiles satisfy equisegmentation property:

$$
R_{ir} = \begin{cases} im + \lfloor u_{ir} \rfloor \text{ if } (r,d) \in A, \text{ and } d \leq 50 & (3.1a) \\ im + \lceil u_{ir} \rceil \text{ if } (r,d) \in A, \text{ and } d > 50 & (3.2b) \\ im + u_{ir} \text{ if } (r,d) \notin A & (3.2c) \end{cases}
$$

where i and r are integers with $1 \leq i \leq 99$, $0 \leq r \leq 99$, $u_{ir} = i(r+1)/100 = \lfloor u_{ir} \rfloor + d/100$ and an admissible set A={(*r*,*d*)} in Table 1 (See Appendix) with $m = (n-r)/100 \geq 1$ observations in each segment.

Example 3.1 Consider Data 5 with sample size $n = 270 = 100(2) + 70$ i.e. $m = 2,\ r = 70$. Using the ranks $R_{50,70}$, $R_{51,70}$, $R_{52,70}$, $R_{53,70}$, $R_{54,70}$, $R_{55,70}$ for the percentiles we have:

$$
\begin{aligned}
u_{50,70} &= 50(70+1)/100 = 35 + 50/100,\ (r,d) = (70,50) \notin A \Rightarrow R_{50,70} = 50(2) + u_{50,70} = 135 + 50/100\\
u_{51,70} &= 51(70+1)/100 = 36 + 21/100,\ (r,d) = (70,21) \in A \Rightarrow R_{51,70} = 51(2) + \lfloor u_{51,70} \rfloor = 138\\
u_{52,70} &= 52(70+1)/100 = 36 + 92/100,\ (r,d) = (70,92) \in A \Rightarrow R_{52,70} = 52(2) + \lceil u_{52,70} \rceil = 141
\end{aligned}
$$

$$u_{53,70} = 53(70+1)/100 = 37 + 63/100,\ (r,d) = (70,63) \notin A \Rightarrow R_{53,70} = 53(2) + u_{53,70} = 143 + 63/100$$
$$u_{54,70} = 54(70+1)/100 = 38 + 34/100,\ (r,d) = (70,34) \in A \Rightarrow R_{54,70} = 54(2) + \lfloor u_{54,70} \rfloor = 146$$
$$u_{55,70} = 55(70+1)/100 = 39 + 5/100,\ (r,d) = (70,5) \in A \Rightarrow R_{55,70} = 55(2) + \lfloor u_{55,70} \rfloor = 149\,.$$

Clearly the ranks satisfy equisegmentation property. The positions of percentiles for Data 5 is given by

$$Q_{50,70} = (50/100)(769) + (50/100)(773) = 771,$$
$$Q_{51,70} = 780\,, \quad Q_{52,70} = 790$$
$$Q_{53,70} = (37/100)(797) + (63/100)(801) = 799.5\,,$$
$$Q_{54,70} = 808\,, \quad Q_{55,70} = 818$$

$$\ldots, x_{(134)}, x_{(135)}, \Downarrow x_{(136)}, x_{(137)}, \overset{\Downarrow}{x}_{(138)}, x_{(139)}, x_{(140)}, \overset{\Downarrow}{x}_{(141)}, x_{(142)},$$
$$x_{(143)}, \Downarrow x_{(144)}, x_{(145)}, \overset{\Downarrow}{x}_{(146)}, x_{(147)}, x_{(148)}, \overset{\Downarrow}{x}_{(149)}, x_{(150)}, x_{(151)}, \ldots$$

where $x_{(i)}$'s are the ordered sample observations. The above rounding of ranks guarantees the desirable equisegmentation property. There are $m = 2$ observations in each segment here. The remainder r (which is 70 here) is also, as expected, the number of percentiles having integer ranks for any sample of size $n \geq 100$.

Acknowledgements

The authors are grateful to King Fahd University of Petroleum and Minerals, Saudi Arabia for excellent research support.

References

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

Joarder, A.H. and Latif, R.M. (2004). A comparison and contrast of some methods for sample quartiles. *Journal of Probability and Statistical Science.* 2(1), 95-109.

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

Mendenhall, W. and Sincich, T. (1995). *Statistics for Engineering and the Sciences*. John Wiley. New York.

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

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

Appendix

r	*d*	*r*	*d*
2	1, 2	31	d=4k,1 ≤ k ≤ 7
5	2, 4	32	1 ≤ d ≤ 32
6	1 ≤ d ≤ 6	33	d=2k,1 ≤ k ≤ 16
7	4	34	d=5k,1 ≤ k ≤ 6
8	1 ≤ d ≤ 8	35	d=4k,1 ≤ k ≤ 8
10	1 ≤ d ≤ 10	36	1 ≤ d ≤ 36
11	4, 8	37	d=2k,1 ≤ k ≤ 18
12	1 ≤ d ≤ 12	38	1 ≤ d ≤ 38
13	d=2k,1 ≤ k ≤ 6	39	20
14	5,10	40	1 ≤ d ≤ 40
15	4, 8, 12	41	d=2k,1 ≤ k ≤ 20
16	1 ≤ d ≤ 16	42	1 ≤ d ≤ 42
17	d=2k,1 ≤ k ≤ 8	43	d=4k,1 ≤ k ≤ 10
18	1 ≤ d ≤ 18	44	d=5k,1 ≤ k ≤ 8
20	1 ≤ d ≤ 20	45	d=2k,1 ≤ k ≤ 22
21	d=2k,1 ≤ k ≤ 10	46	1 ≤ d ≤ 46
22	1 ≤ d ≤ 22	47	d=4k,1 ≤ k ≤ 11
23	d=4k,1 ≤ k ≤ 5	48	1 ≤ d ≤ 48
25	d=2k,1 ≤ k ≤ 12	50	1 ≤ d ≤ 49, 99
26	1 ≤ d ≤ 26	51	d=4k,1 ≤ k ≤ 12
27	d=4k,1 ≤ k ≤ 6	52	1 ≤ d ≤ 49, 97 ≤ d ≤ 99
28	1 ≤ d ≤ 28	53	d=2k, 1 ≤ k ≤ 24, 48 ≤ k ≤ 49
29	10, 20	54	d=5k, 1 ≤ k ≤ 9, 95
30	1 ≤ d ≤ 30	55	d=4k, 1 ≤ k ≤ 12, 96

Table 1a: set A = {(*r*,*d*)} for Percentile (*f* = 100)

r	d	r	d
56	1 ≤ d ≤ 49, 93 ≤ d ≤ 99	78	1 ≤ d ≤ 49, 71 ≤ d ≤ 99
57	d=2k, 1 ≤ k ≤ 24, 46 ≤ k ≤ 49	79	20,40,80
58	1 ≤ d ≤ 49, 91 ≤ d ≤ 99	80	1 ≤ d ≤ 49, 69 ≤ d ≤ 99
59	20, 40	81	d=2k, 1 ≤ k ≤ 24, 34 ≤ k ≤ 49
60	1 ≤ d ≤ 49, 89 ≤ d ≤ 99	82	1 ≤ d ≤ 49, 67 ≤ d ≤ 99
61	d=2k, 1 ≤ k ≤ 24, 44 ≤ k ≤ 49	83	d=4k, 1 ≤ k ≤ 12, 17 ≤ k ≤ 24
62	1 ≤ d ≤ 49, 87 ≤ d ≤ 99	84	d=5k, 1 ≤ k ≤ 9, 13 ≤ k ≤ 19
63	d=4k, 1 ≤ k ≤ 12, 22 ≤ k ≤ 24	85	d=2k, 1 ≤ k ≤ 24, 32 ≤ k ≤ 49
64	d=5k, 1 ≤ k ≤ 9, 17 ≤ k ≤ 19	86	1 ≤ d ≤ 49, 63 ≤ d ≤ 99
65	d=2k, 1 ≤ k ≤ 24, 42 ≤ k ≤ 49	87	d=4k, 1 ≤ k ≤ 12, 16 ≤ k ≤ 24
66	1 ≤ d ≤ 49, 83 ≤ d ≤ 99	88	1 ≤ d ≤ 49, 61 ≤ d ≤ 99
67	d=4k, 1 ≤ k ≤ 12, 21 ≤ k ≤ 24	89	d=10k, 1 ≤ k ≤ 9, d ≠ 50
68	1 ≤ d ≤ 49, 81 ≤ d ≤ 99	90	1 ≤ d ≤ 49, 59 ≤ d ≤ 99
69	d=10k, 1 ≤ k ≤ 9, d ≠ 50	91	d=4k, 1 ≤ k ≤ 12, 15 ≤ k ≤ 24
70	1 ≤ d ≤ 49, 79 ≤ d ≤ 99	92	1 ≤ d ≤ 49, 57 ≤ d ≤ 99
71	d=4k, 1 ≤ k ≤ 12, 20 ≤ k ≤ 24	93	d=2k, 1 ≤ k ≤ 24, 28 ≤ k ≤ 49
72	1 ≤ d ≤ 49, 77 ≤ d ≤ 99	94	d=5k, 1 ≤ k ≤ 9, 11 ≤ k ≤ 19
73	d=2k, 1 ≤ k ≤ 24, 38 ≤ k ≤ 49	95	d=4k, 1 ≤ k ≤ 12, 14 ≤ k ≤ 24
74	25, 75	96	1 ≤ d ≤ 49, 53 ≤ d ≤ 99
75	d=4k, 1 ≤ k ≤ 12, 19 ≤ k ≤ 24	97	d=2k, 1 ≤ k ≤ 24, 26 ≤ k ≤ 49
76	1 ≤ d ≤ 49, 73 ≤ d ≤ 99	98	1 ≤ d ≤ 49, 51 ≤ d ≤ 99
77	d=2k,1 ≤ k ≤ 24, 36 ≤ k ≤ 49		

Table 1b: set A = $\{(r,d)\}$ for Percentile ($f = 100$)

650	762	797	468	1193	787	738	1098	475	360
720	412	1112	1210	556	689	1189	1011	591	503
902	566	1000	958	384	885	388	811	780	1235
979	941	916	1214	1077	1060	339	976	346	363
1039	906	755	1133	587	311	444	808	1105	923
395	731	1238	846	1158	1007	1179	1018	636	871
419	962	472	878	972	370	741	794	570	325
685	615	818	1200	934	955	1035	454	668	909
608	633	605	1046	920	748	598	398	314	1186
510	1063	776	542	374	307	790	997	507	500
1130	769	804	381	1074	300	318	356	517	1109
353	612	493	993	986	675	892	724	521	433
647	752	304	836	377	1168	349	843	601	409
745	1154	496	1207	1091	328	734	482	437	930
965	416	573	489	1025	1172	1123	860	1053	538
650	762	797	468	1193	787	738	1098	475	360
720	412	1112	1210	556	689	1189	1011	591	503
902	566	1000	958	384	885	388	811	780	1235
979	941	916	1214	1077	1060	339	976	346	363
1039	906	755	1133	587	311	444	808	1105	923
395	731	1238	846	1158	1007	1179	1018	636	871

419	962	472	878	972	370	741	794	570	325
685	615	818	1200	934	955	1035	454	668	909
608	633	605	1046	920	748	598	398	314	1186
510	1063	776	542	374	307	790	997	507	500
1130	769	804	381	1074	300	318	356	517	1109
353	612	493	993	986	675	892	724	521	433

Data 5