☒Teacher's Coner

4분위수에 대한 메모

박동준 · 황현미
부경대학교 수리과학부

A Note on Quartile

Dong Joon Park · Hyun Mi Hwang
Division of Mathematical Sciences, Pukyong National University

Abstract

It is necessary to describe a data set after collection of data in elementary statistics course. Two major numerical summary of the data set may be measures of central location and dispersion. There are various numerical summary methods in presenting how data are dispersed and each method has its own advantages and disadvantages. Quartiles are discussed among several methods to describe dispersion of data set. When data type is discrete, exact quartile values are sometimes ambiguous to find, whereas exact quartile values are obtained for continuous data. Examples of both data types are given. Programs listed below may be used to provide quartiles in MINITAB and SAS.

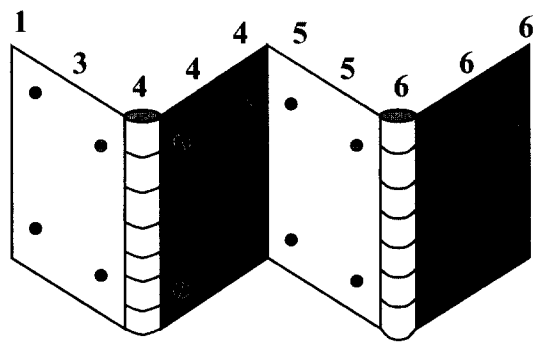
1. Discrete data

The concept of a central tendency is introduced to describe data set. The measures of central location often used are mean, median, and mode. The measures of dispersion of a data set are range, percentile, decile, quartile, interquartile range, variance, standard deviation, coefficient of variation and so on. Median tells us the location that divides the data set half. When individual measurements of a data set are arranged in ascending order, a rudimentary definition of median is the middle value if the number of observations is odd and median is the average of two middle values if the number of observations is even.

Quartiles mean the position that divides the halves divided by median again. That is, quartiles are the positions that break the ordered statistics into four pieces. Thus, there are three quartiles and they are also called hinges. Consider following ten observations:

1, 3, 4, 4, 4, 5, 5, 6, 6, 6.

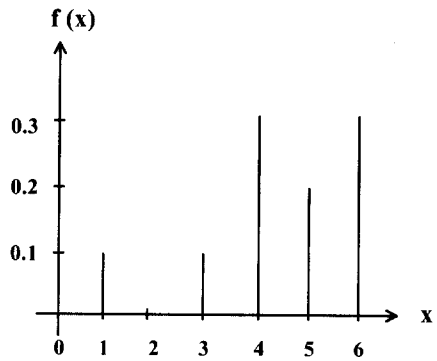
By arranging observations on two hinges put in parallel in <Figure 1>, the quartiles are 4, 4.5, and 6.



< Fig 1 > Numbers arranged on hinges

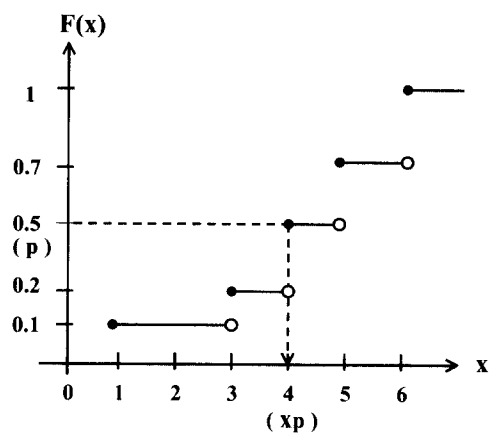
If the number of observations is odd, the middle value should be included to find the first and third quartiles. It is sometimes ambiguous to find exact values for quartiles when data is discrete.

Smaller division of a data set proceeds to percentiles. A percentile is a number below which a specific percent of a data set lies. Hogg and Craig(1995) discussed that a (100p)th percentile of the distribution of a random variable X is a value X_p such that $\Pr(X < X_p) \leq p$ and $\Pr(X \leq X_p) \geq p$ where $0 < p < 1$. Three quartiles are 25^{th}, 50^{th}, and 75^{th} percentiles. The 25^{th} percentile is called the first quartile or lower quartile, the 50^{th} percentile is called the second quartile, middle quartile, or median, and the 75^{th} percentile is called the third quartile or upper quartile. When the probability mass function (p.m.f.) for a data set is given in the form of table, equation, or graph, cumulative distribution function(c.d.f.) can be used to apply Hogg and Craig's definition. For example, p.m.f. of ten measurements of the data set is given in <Figure 2>.



< Fig 2 > p.m.f. of data set

The c.d.f. of the data set is drawn in <Figure 3> and the first quartile is 4, median is all numbers in interval from 4 to 5, and third quartile is 6.



< Fig 3 > c.d.f. of data set

On the other hand, statistical software packages have their own methods to find quartiles. MINITAB uses linear interpolation. After inputting ten observations in spreadsheet, we type DESCRIBE at MTB prompt. The results of descriptive statistics are then given as below.

[MINITAB PROGRAM]
MTB > DESCRIBE C1
Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev
C1	10	4.400	4.500	4.625	1.578
SEMean		Min	Max	Q1	Q3
0.499		1.000	6.000	3.750	6.000

SAS provides five different methods to find quartiles. SAS provides 3.5, 3, 4, 3.75, and 4 as the first quartile for method 1, 2, 3, 4, and 5, respectively, by specifying number from 1 to 5 in option PCTLDEF following procedure PROC UNIVARIATE. The necessary codes of finding descriptive statistics in SAS are as follows:

```
[SAS PROGRAM]
DATA A;
INPUT SCORE @@;
CARDS;
1 3 4 4 4 5 5 6 6 6
;
PROC UNIVARIATE PCTLDEF = 5;
VAR SCORE;
RUN;
```

Quantiles(Def=5)

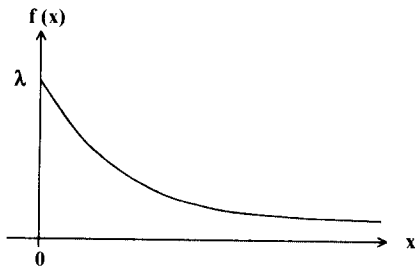
100%	Max	6	99%	6
75%	Q3	6	95%	6
50%	Med	4.5	90%	6
25%	Q1	4	10%	2
0%	Min	1	5%	1
			1%	1.

2. Continuous data

If data are continuous, quartiles are clearly found once the c.d.f. of continuous data is obtained. Exponential distribution is illustrated to find the first quartile. The probability density function(p.d.f.) is denoted by

$$f(x)=\lambda e^{-\lambda x} \quad ; \quad x \geq 0$$

and its graph is drawn in <Figure 4>.

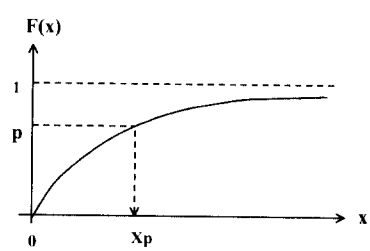


< Fig 4 > p.d.f. of exponential distribution

The c.d.f. of exponential distributions is then written as

$$F(x)=1-e^{-\lambda x} \quad ; \quad x \geq 0$$

and its c.d.f. is drawn in <Figure 5>.



< Fig 5 > c.d.f. of exponential distribution

The first quartile, $X_{0.25}$, is obtained by solving $F(X_{0.25})=0.25$ with respect to $X_{0.25}$, i.e.,

$$X_{0.25} = \ln 0.75 / -\lambda.$$

3. Conclusion

Quartiles, one of measures of dispersion, were discussed. Unlike other measures exact quartile values are ambiguous to find when data type is discrete, a rudimentary way for obtaining quartiles was suggested drawing hinges. More precise method for finding quartiles is provided by citing Hogg and Craig's definition. One the other hand, exact quartile values were obtained by developing c.d.f.

Reference

[1] Hogg, R.V. and Craig, A.T(1995), *Introduction to Mathematical Statistics*, the 5^{th} edition, Prentice Hall