Quartile

From Wikipedia, the free encyclopedia

In descriptive statistics, a **quartile** is any of the three values which divide the sorted data set into four equal parts, so that each part represents one fourth of the sampled population.

In epidemiology, the **quartiles** are the four ranges defined by the three values discussed here.

Contents

Definitions

- **first quartile** (designated Q_1) = **lower quartile** = cuts off lowest 25% of data = *25th percentile*
- **second quartile** (designated Q_2) = *median* = cuts data set in half = *50th percentile*
- **third quartile** (designated Q_3) = **upper quartile** = cuts off highest 25% of data, or lowest 75% = *75th percentile*

The difference between the upper and lower quartiles is called the *interquartile range*.

Computing methods

There is no universal agreement on choosing the quartile values.[1]

The formula for locating the position of the observation at a given percentile, *y*, with *n* data points sorted in ascending order is:[1]

$$L_y = (n)\left(\frac{y}{100}\right)$$

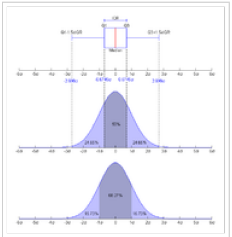
- Case 1: If L is a whole number, then the value will be found halfway between positions L and L+1
- Case 2: If L is a decimal, round up to the nearest whole number. (for example, L = 1.2 becomes 2)

Example

One possible rule (employed by the TI-83 calculator boxplot and 1-Var Stats functions) is as follows:

1. Use the median to divide the ordered data set into two halves. Do not include the median into the halves.
2. The lower quartile value is the median of the lower half of the data. The upper quartile value is the median of the upper half of the data.

The examples below assume this rule. Another possible rule would be to include the median in the halves when calculating the quartiles. This would give significantly different answers to the examples.



Example 4. Boxplot (with quartiles and an interquartile range) and a probability density function (pdf) of a normal $N(0,1\sigma^2)$ population

Example 1
Data Set: 6, 47, 49, 15, 42, 41, 7, 39, 43, 40, 36
Ordered Data Set: 6, 7, 15, 36, 39, 40, 41, 42, 43, 47, 49

$$\begin{cases} Q_1 = 15 \\ Q_2 = 40 \\ Q_3 = 43 \end{cases}$$

Example 2
Ordered Data Set: 7, 15, 36, 39, 40, 41

$$\begin{cases} Q_1 = 15 \\ Q_2 = 37.5 \\ Q_3 = 40 \end{cases}$$

Example 3
Ordered Data Set: 1, 2, 3, 4

$$\begin{cases} Q_1 = 1.5 \\ Q_2 = 2.5 \\ Q_3 = 3.5 \end{cases}$$

In line with the definition of the median (given on the Wikipedia page for Median) as; P(X<=m)>=0.5 and P(X>=m)>=0.5, it seems sensible to define the lower quartile, l, as P(X<=l)>=0.25 and P(X>=l)>=0.75. This means that for;

```
4n-3 pieces of data, the LQ is the n th value,
4n-2 pieces of data, the LQ is the n th value,
4n-1 pieces of data, the LQ is the n th value,
```

but for 4n pieces of data, the LQ is not uniquely defined, by this, it could be the n th value, or the n+1 th value. It would be sensible to take the average of these two values in this case. The Upper Quartile, u, is similarly defined a; P(X<=u)>=0.75 and P(X>=u)>=0.25. It is then easier to count the same number, from the RIGHT.

See also

- Range
- Quantile
- Quarter
- Summary statistics

External links

- Quartile - from MathWorld Includes references and compares various methods to compute quartiles
- Quartiles - From MathForum.org
- Quartiles - An example how to calculate it

References

1. ^ http://books.google.com/books?id=1LH6tNn6CYkC&printsec=frontcover&source=bl&ots=lOg76JIira&sig=Jp_OJYojlBs0LszvhIKuWkEjBuM&hl=en&ei=U4NdSszRLoqGsgPywYCxCg&sa=X&oi=book_result&ct=r