Quartiles, quintiles, centiles, and other quantiles

Douglas G Altman, J Martin Bland

When presenting or analysing measurements of a continuous variable it is sometimes helpful to group subjects into several equal groups. For example, to create four equal groups we need the values that split the data such that 25% of the observations are in each group. The cut off points are called quartiles, and there are three of them (the middle one also being called the median). Likewise, we use two tertiles to split data into three groups, four quintiles to split them into five groups, and so on. The general term for such cut off points is quantiles; other values likely to be encountered are deciles, which split data into 10 parts, and percentiles, which split the data into 100 parts (also called percentiles). Values such as quartiles can also be expressed as centiles; for example, the lowest quartile is also the 25th centile and the median is the 50th centile. We consider below some common applications of quantiles.

A common confusion is to use the terms tertiles, quartiles, quintiles, etc, not for the cut off points but for the groups so obtained, but these are properly called thirds, quarters, fifths, and so on.

**Data description**—The mean and standard deviation are useful to summarise a set of observations. When the data have a skewed distribution it is often preferable to quote instead the median and two outer centiles, such as the 10th and 90th. The first and third quartiles (25th and 75th centiles) are sometimes used; these define the interquartile range. The median is a useful summary statistic when some of the values are not actually measured—for example, because some values are outside the range of the measuring equipment. Similarly, the median is frequently used when summarising survival data, when it is usual for some of the survival times to be unknown.

**Reference intervals and centiles**—A special type of data description arises in the construction of a reference interval (normal range). A 95% reference interval is defined by the values that cut off 2 1/2% at each end of the distribution. (These values are often quite reasonably called the 2 1/2% and 97 1/2% centiles, although it is not strictly correct to have half centiles.) Reference intervals are widely used in clinical chemistry. By contrast, charts for the assessment of human size or growth usually show several centiles. Reference centiles are sometimes defined using the normal distribution, which is used in such cases. The many groups created the greater is the loss of information. In regression analyses continuous explanatory variables are often categorised into two or more groups. Although this slightly complicates the analysis, it avoids a direct assumption that there is a linear relation between the variable and the outcome of interest. However, it leads to a model in which risk apparently jumps at certain values of the predictor variable rather than increasing smoothly.

**Calculation of quantiles**—The calculation of centiles and other quantiles is not as simple as it might seem. The data should be ranked from 1 to n in order of increasing size. The kth centile is obtained by calculating \( q = k(n+1)/100 \) and then interpolating between the two values with ranks either side of the \( q \)th. For example, for the 5th centile of a sample of 145 observations we have \( q = 5 \times 146/100 = 7.3 \). We estimate the 5th centile as the value 0.3 of the way between the 7th and 8th ranked observations. If these data values are 11-4 and 14-9 the estimated centile is 12.45. Confidence intervals can be constructed for any quantile.}

---

**Notes**