Computing Percentiles

Suppose that we have a data set $X_1, \ldots, X_n$. The $p$th percentile is conceptually meant to be a value $Q_p$ such that $p$ proportion (or $100p\%$) of the observations fall below $Q_p$. It so happens that there are different ideas about how to compute $Q_p$ in practice. See, for example,

http://en.wikipedia.org/wiki/Percentile

for useful discussion.

**The Method used by NIST**

The National Institute of Standards and Technology (NIST) uses the following approach:

$$Q_p = \text{the observation of rank } (n + 1)p \text{ (obtained by interpolation if necessary).}$$

For example, *Microsoft Excel* and *Minitab* use this approach.

**Illustration**

Consider the data set $X_1, \ldots, X_n = \{39, 45, 46, 9, 16, 7, 48, 51, 45, 36\}$. Here $n = 10$. The ordered values are

$$7 \ 9 \ 16 \ 36 \ 39 \ 45 \ 45 \ 46 \ 48 \ 51$$

Let us find the 0.5 or 50th percentile (i.e., the median). By the *NIST* method, we seek the observation of rank $(n + 1)p = 11 \times 0.5 = 5.5$. Thus we go halfway between the 5th and 6th observations, which means that we average these two observations. Hence the median is

$$Q_{0.5} = \frac{39 + 45}{2} = 42.0.$$

Now let us find the 1st quartile $Q_{0.25}$. Using the *NIST* method, we seek the observation of rank $(n + 1)p = 11 \times 0.25 = 2.75$. This is again obtained by interpolation: We get the 2nd observation and add to it 0.75 times the distance between it and the 3rd observation. That is,

$$Q_{0.25} = 9 + (0.75) \times (16 - 9) = 9 + 5.25 = 14.25.$$

For the 3rd quartile, or 75th percentile, that is, for $Q_{0.75}$, check that the *NIST* method gives $46.5$.

– RJS, 2/12/2014