

## 6.1: Discrete Data Percentiles and Quartiles

Before we get into how to calculate percentile in a data set, note that we can see percentiles directly on a cumulative frequency plot, see Figure 6.2.

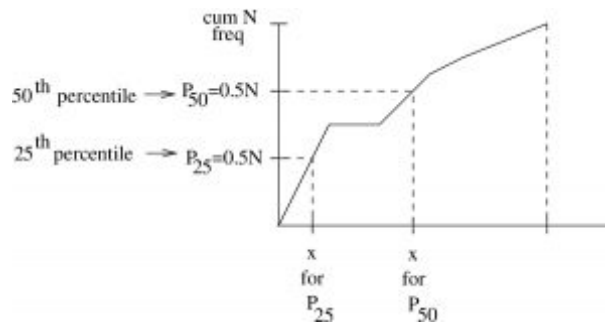


Figure 6.2 : With a cumulative frequency plot, we can read percentiles off the  $y$  axis. If you have a newborn baby and take it to the doctor for their first check up, they will measure the baby's head circumference and tell you the baby's head size percentile by looking at such a chart. The doctor's chart will be based on an accumulation of a very large number of essentially population data. Cumulative frequency graphs, or more exactly cumulative probability graphs, can be made for continuous distributions like the normal distribution. The resulting function is the Cumulative Distribution Function, or CDF, and is, for example,  $P(z)$  represents the  $z$ -distribution then  $CDF(x) = \int_0^x P(z)dz$ . We will see this CDF in SPSS.

*Computing percentile positions of discrete data.* Let  $i$  be the ordered position of a data set of  $n$  data points, then we define the percentile position of  $x_i$  to be

$$P(x_i) = \frac{(i-1)}{(n-1)} \times 100.$$

This formula has the property that  $P(x_1 = L) = 0$  and  $P(x_n = H) = 100$ . It is what we will use as a percentile formula but it is not the only one. Look at Figure 6.1. The way the histogram there is shaded the formula would be  $P(x_i) = \frac{i}{n} \times 100$  which would have the property that  $P(L) = \frac{100}{n}$  and  $P(H) = 100$ . There are other, not necessarily wrong, ways to define the percentile position of discrete data but we will use Equation 6.1.

If you want to find the position,  $i$ , of the data point corresponding to a given percentile  $P$  then compute

$$i = \left\lceil \frac{P \times (n-1)}{100} \right\rceil + 1.$$

Equation (6.2) is derived by solving Equation (6.1) for  $i$ . Note that Equation (6.2) gives the *position* of the data point  $x_i$ , not its value. To clarify that, let's look at an example.

**Example 6.1 :** Consider the dataset given below. Data would originally be given as the numbers in the first line. So the first step in answering any question about percentiles is to order the data, the same as what you need to to determine the median of a dataset. Once the data are ordered, then you may assign a position number to each data point as shown in the third line.

original data	18	15	12	6	8	2	3	5	20	10
ordered data	2	3	5	6	8	10	12	15	18	20
$i$	1	2	3	4	5	6	7	8	9	10
$n = 10$										

Q : What is the percentile rank of  $x_i = 12$ ?

A :  $i = 7$  so  $P(12) = P(x_7) = \frac{(7-1)}{10-1} \times 100 = \frac{6}{9} \times 100 = 67^{\text{th}}$  percentile.

Q : What is the value corresponding to the 25<sup>th</sup> percentile,  $P_{25}$ ?

A :  $i = \left\lceil \frac{P \times (n-1)}{100} \right\rceil + 1 = \left\lceil \frac{(25) \times (10-1)}{100} \right\rceil + 1 = \left\lceil \frac{25 \times 9}{100} \right\rceil + 1 = 2.25 + 1 = 3.25$

The closest  $i$  is 3 and  $x_3 = 5$ . We can write  $P_{25} = 5$ .

**Decile :**

$D(x_i) \equiv$  The decile of data value  $x_i$  in the *ordered* position  $i$  is defined as

$$D(x_i) = \frac{P(x_i)}{10} \quad 0 \leq D(x_i) \leq 10 \quad (6.1.1)$$

We will not make much use of decile except to see that quartile is defined in the same way.

**Quartile :**

$Q(x_i) \equiv$  The quartile of data value  $x_i$  in the ordered position  $i$ .

$$Q(x_i) = \frac{P(x_i)}{25} \quad 0 \leq Q(x_i) \leq 4$$

*Notation :* (This notation also applies to  $P$  and  $D$ .) We write :

$Q_0$  &= & 0<sup>th</sup> quartile

$Q_1$  &= & 1<sup>st</sup> quartile

$Q_2$  &= & 2<sup>nd</sup> quartile

$Q_3$  &= & 3<sup>rd</sup> quartile

$Q_4$  &= & 4<sup>th</sup> quartile

Quartiles are useful because we do not have to compute percentile first and then divide by 25 as given by Equation (6.3). Instead, we can use the following handy tricks after ordering our data:

$$\begin{aligned} Q_2 &= \text{MD (median)} \\ Q_1 &= \text{MD of values less than } Q_2 \\ Q_3 &= \text{MD of values greater than } Q_2 \\ Q_0 &= L \\ Q_4 &= H \end{aligned}$$

**Example 6.2 :** Example with an even number of data points. With the data *in order*, first find the median, then the medians of the two halves of the dataset :

$$5 \quad 6 \quad 12 \quad 13 \quad 15 \quad 18 \quad 22 \quad 50 \quad (6.1.2)$$

$$Q_1 = \frac{6+12}{2} = 9$$

$$MD = \frac{13+15}{2} = 14 = Q_2$$

$$Q_3 = \frac{18+22}{2} = 20$$

$$Q_0 = L = 5$$

$$Q_4 = H = 50$$

**Example 6.3 :** Example with an even number of data points. With the data *in order*, first find the median, then the medians of the two halves of the dataset :

$$2 \quad 5 \quad 11 \quad 14 \quad 18 \quad 25 \quad 35 \quad (6.1.3)$$

$$Q_1 = 5$$

$$MD = 14 = Q_2$$

$$Q_3 = 25$$

$$Q_0 = L = 2$$

$$Q_4 = H = 35$$

This page titled [6.1: Discrete Data Percentiles and Quartiles](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Gordon E. Sarty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.