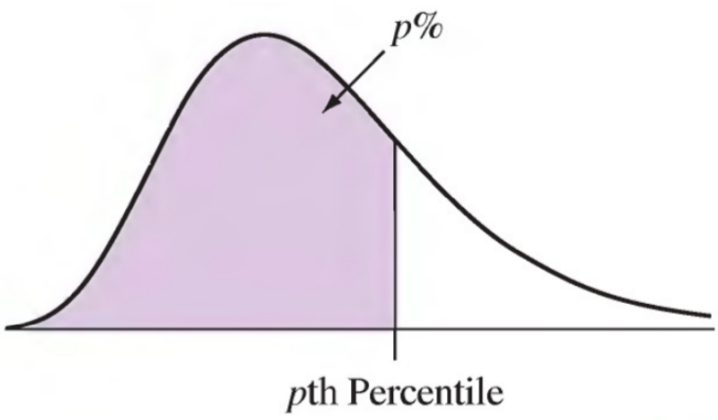
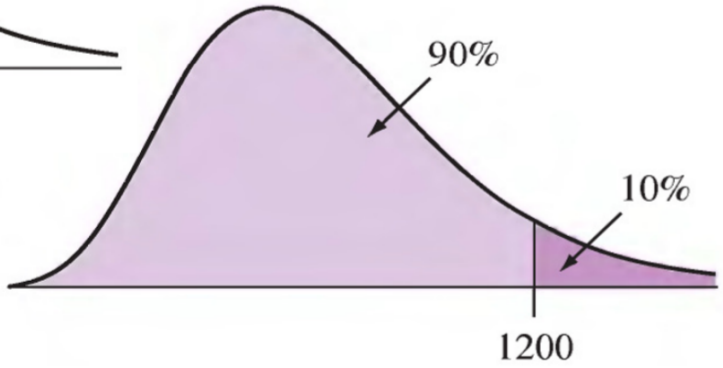
Measures of Position

Percentile

Percentile The p^{th} percentile is a value such that p % of the observations fall below or at that value.

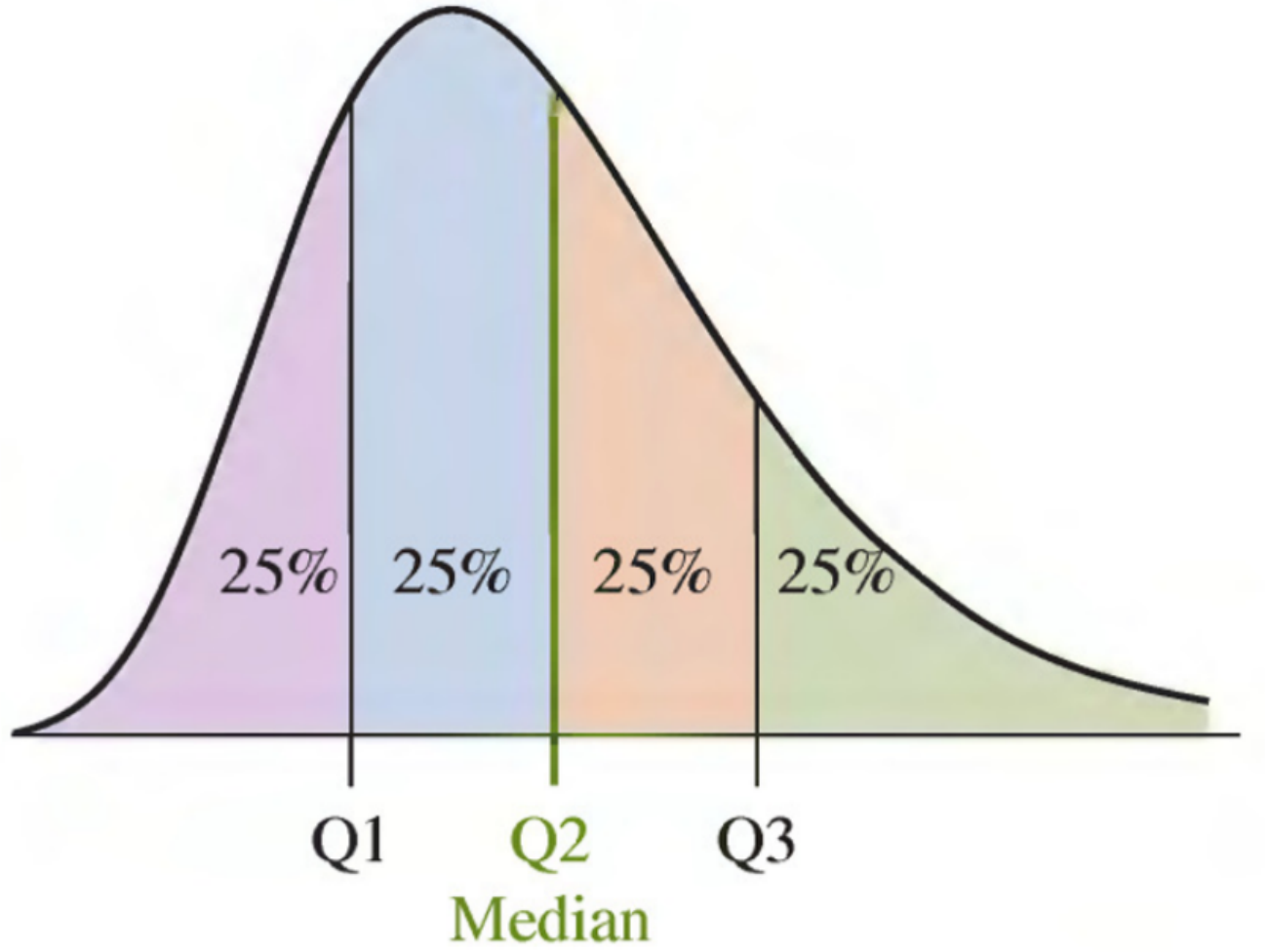




Quartiles

The Quartiles split the distribution into four parts that have the same number of observations:

- 25% is below the first quartile (Q1),
- 25% is between the first quartile and the second quartile (the median, Q2),
- 25% is between the second quartile and the third quartile (Q3), and
- 25% is above the third quartile.



The procedure to find them is:

- Arrange the data in order.
- Consider the median. This is the second quartile, Q2.
- Consider the lower half of the observations (excluding the median itself if n is odd). The median of these observations is the first quartile, Q1.
- Consider the upper half of the observations (excluding the median itself if n is odd). Their median is the third quartile, Q3.

Important

It is not the range that we are dividing, but the observations

In fact, not all quartiles are equal in x range.

Interquartile Range (IQR)

It is the distance between the third quartile and first quartile.

$$Q3 - Q3$$

Info

Useful for detecting outliers.

In fact, an observation is a potential outlier if it falls more than 1.5×IQR below the first quartile or more than 1.5×IQR above the third quartile.

Example: Cereal Sodium Data

- For the breakfast cereal sodium data has QI=135 and Q3=205. So, IQR = Q3-QI=205-135=70.
- For those data

  $$1.5\times\text{IQR}=1.5\times 70=105.$$

  QI-105=30 (lower boundary, potential outliers below), and Q3+105=310 (upper boundary, potential outliers above).
- By the 1.5×IQR criterion, observations below 30 or above 310 are potential outliers.
- The only observations below 30 or above 310 are the sodium values of 0 and 340 mg. These are the only potential outliers.