Measures of Relative Position

A. Recall Standard Deviation

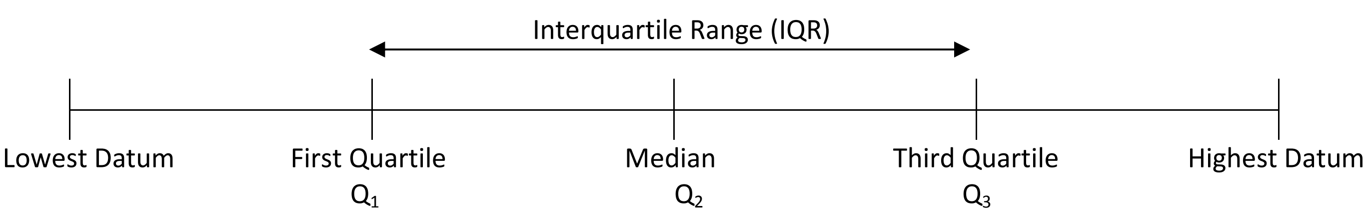
Last class, we began our discussion on *dispersion* by introducing variance and standard deviation:

Population Standard Deviation	Sample Standard Deviation
$\sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{N}}$	$s = \sqrt{\dfrac{\sum(x-\overline{x})^2}{n-1}}$
Population Standard Deviation (Grouped)	Sample Standard Deviation (Grouped)
$\sigma \approx \sqrt{\dfrac{\sum f_i(m_i-\mu)^2}{N}}$	$s \approx \sqrt{\dfrac{\sum f_i(m_i-\overline{x})^2}{n-1}}$

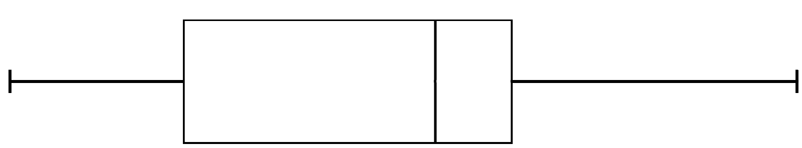
There are two other important measures of dispersion, called **measures of relative position**, which describe the portion of data below a certain data point.

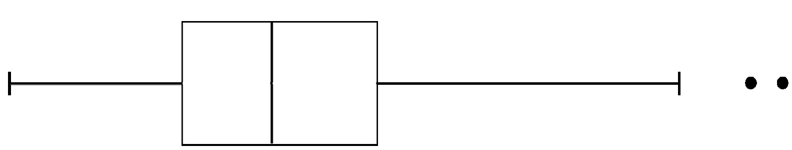
B. Quartiles and Interquartile Range

- The ____________ divide a set of *ordered data* into four equal segments after it has been arranged in ascending order. (This is similar to how the median divides data in two equally sized groups!)
- The _____________________________________(IQR) is the range of the middle half of the data. It has a value of $Q_3 - Q_1$. The larger the interquartile range, the larger the spread. *Note*: the **semi-interquartile range** is one half of the IQR.



A **box-and-whisker plot** and a **modified box-and-whisker plot** illustrate quartiles and interquartile ranges. The only difference between the two is that a *modified* box-and-whisker plot shows *outliers*. If a point is *outside* of $\{Q_1 - 1.5(IQR)\}$ or $\{Q_3 + 1.5(IQR)\}$, it is considered an outlier.





Ex. 1: A random survey of 15 people walking into the school were asked how many times they have attended a live concert. The results were as follows.

3	2	1	10	4	7	35	12
0	1	4	4	3	1	8	

Determine the median, the first and third quartiles, the interquartile range, and any outliers. Draw a modified box-and-whisker plot.

C. Percentiles

Percentiles are similar to quartiles, except that they divide the sets of data into 100 intervals with equal numbers of values. Percentiles bring some confusing notation with them, so we need to define a few variables:

- ✓ Percentiles are labeled as P_k. As such, P_k is an actual item of data. This value is called the k^{th} percentile.
- ✓ In any particular set of data, k% of the data is less than or equal to the value P_k.
- ✓ **<u>ALWAYS</u>** place the data in numerical order when working with percentiles.

For example, P_{80} means that 80% of the data is *less than or equal* to the value of P_{80} and 20% of the data is *greater than or equal to* the value of P_{80}. To find the value of P_{80} first multiply **0.8** by ***n***: this datum and the one above it contain *between them* the 80^{th} percentile. Thus, it is the *average* of these two pieces of data.

Ex. 2: The given set of data summarizes exam scores for one section of STAT 101 at the University of Waterloo.

34	50	58	62	65	68	72	78	84	91
41	51	58	62	65	68	73	79	86	92
45	53	60	63	67	69	75	82	87	96
48	56	62	64	67	70	76	82	89	99

a) Find the 90^{th} percentile.

b) Does a certain student's score of 75% place the student at the 70^{th} percentile?

CW: p. 148 #2 – 4, 6bcd, 7, 14