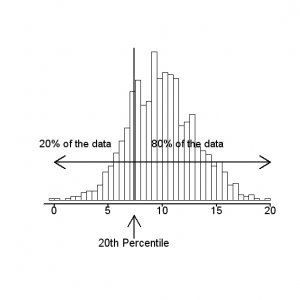
We know now that summary statistics such as mean and standard deviation provide a quick and easy way to analyze probability distributions. However, there are some questions that do not need these statistics to be answered. How did my test scores compare to others? Am I using too much of my utilities? We could answer these using said summary statistics, but percentile ranks are by far a much quicker method. I will be discussing the definition of percentiles and percentile ranks and some examples of their use.

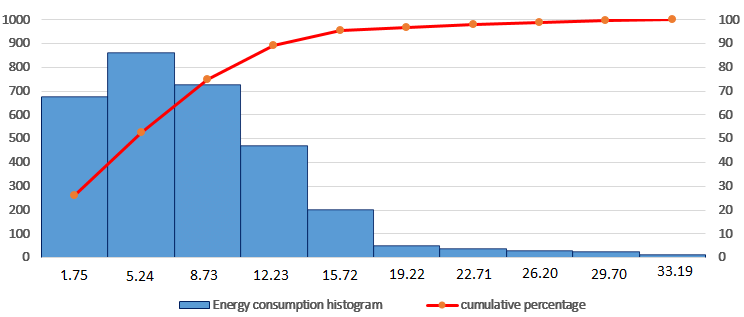
A percentile, as the name suggests, is one of 100 groups that a distribution of data points can be divided into. Their numbering increases from left to right and visualize how many points in a distribution fall at or below a chosen point. Imagine a data set is split into 100 groups where each group contains the same number of points. A point in the 70th percentile means that 70% of the data points fall at or below that point and 30% are above. A percentile *rank* is an *integer* that indicates what percentile the data point is located in for a distribution (i.e., a rank of 70 means a data point is in the 70th percentile). The rank is 100 times the ratio of the number of points at or below a chosen point divided by the total number of points.

A percentile rank is *not* a percent value in the traditional sense. For example, if a dataset of some percentages had the points [15, 15, 30, 30, 45, 60, 75, 90, 90, 100], the data point 30 indicates 30 percent but would have a percentile rank of 40 since 4 of 10 points are equal to or less than 30. The 50th percentile, the average of 45 and 60 here, is better known as the median. The 25th, 50th, and 75th percentiles are known as quartiles due to how they split the data. Using the median and quartiles, the central tendency of the distribution can be determined. Below is a histogram with the 20th percentile highlighted, and we can see how the data is spread out and where a majority of the points lie.



When would we actually use percentiles? Comparing test scores is by far the most common use. A test taker is curious to know how many people they scored better than, and percentiles reveal this exactly. If I looked at a histogram of scores with the data divided into 100 groups, I would know immediately how many people I did better than. Percentiles can also be used to objectively compare a data point across two distributions with different means and standard deviations, similar to the usage of Z-scores. If I wanted to see how I would have fared in a different testing year, I could find what percentile my score would have been in that year’s distribution.

We can use percentile ranks to determine a household’s utility usage across communities or times. Let’s say we have a distribution of the monthly electricity usage for all houses in a community (see the example below). The percentile rank is useful in determining what percentile a household is in, or how many houses use less energy. This can be a useful tool for companies to encourage energy conservation. PG&E, the utility provider for much of where I live, frequently (but not always) compares a house’s monthly energy usage to other houses within a distance. Assessing percentiles rather than ranks is much harder for this example, as a sample size of households is much greater than that of test scores.



It is important to note that the terms *percentile* and *percentile rank* cannot be used interchangeably, despite the latter representing the former. Percentiles are groups splitting data that are described by the percentile rank.

References:

Corso, T. & Cerquitelli, T. (2018). *Meteorological Data Analysis for Thermal Energy Characterization by Means of Self-Learning Transparent Models.* ResearchGate. Retrieved on September 20, 2020 from https://www.researchgate.net/figure/Histogram-and-cumulative-distribution-of-thermal-energy-consumption\_fig3\_325356932.

Downey, A. B. (2014). *Think Stats.* O’Reilly Media.

Glen, S. (2020). *Percentiles, Percentile Rank & Percentile Range: Definition & Examples.* Statistics HowTo. Retrieved on September 20, 2020 from https://www.statisticshowto.com/probability-and-statistics/percentiles-rank-range/.

Rumsey, D. J. (2020). *What Percentile Tells You about a Statistical Value.* Dummies by Wiley. Retrieved on September 20, 2020 from https://www.dummies.com/education/math/statistics/what-percentile-tells-you-about-a-statistical-value/.