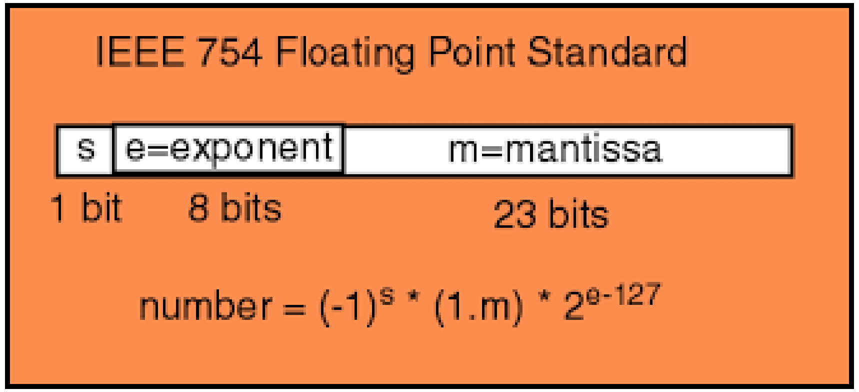
Digital Representation and Computational Errors

Alternatively, errors are introduced during arithmetic operations due to the manner in which numbers are stored on a digital computer (**roundoff error**), or due to the fact that computers can only deal with discrete values of continuous functions (**truncation error**).

When reporting results, numbers should be given to no more significant figures than one beyond that representing the precision of the data.

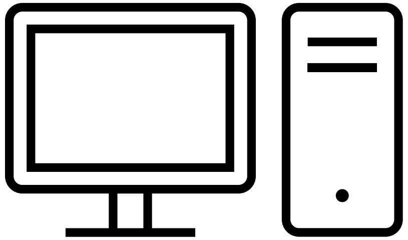
Roundoff error: arises when the amount of storage available for the number is not large enough to accommodate the full numerical information, so the computer simply rounds it off to a size it can store. For example, a typical (IEEE standard) storage scheme for a floating-point number uses 32 bits, of which 1 bit is used to specify the sign of the number, 8 bits to define the exponent, and 23 bits to define the mantissa.



Digital Representation and Computational Errors

One must try to minimize roundoff errors as much as possible. Strategies can be increasing the storage capacity with double-precision floating-point numbers or altering the manner in which certain calculations are done, such as by subtracting off a large constant before summing numbers that differ by a small amount.

Truncation error: arises when the discrete approximation to a continuous function is computed using a limited number of terms in an infinite expansion or using a limited number of points to represent the function. This error can be minimized through the use of many algorithms that we are not going to detail at this stage.



Noise

One of the goals of data analysis is to detect meaningful signals/patterns in **noise** or reduce the degree of noise if possible.

Noise can be classified according to its contribution relative to some more stable (non-fluctuating) component of the observations referred to as the **signal**.

Signal-to-Noise Ratio (SN) scale in decibels (dB):

$$SN = 10\ \log_{10}\left(\frac{power\ of\ signal}{power\ of\ noise}\right)$$

or

$$SN = 20\ \log_{10}\left(\frac{|amplitude\ of\ signal|}{|amplitude\ of\ noise|}\right)$$

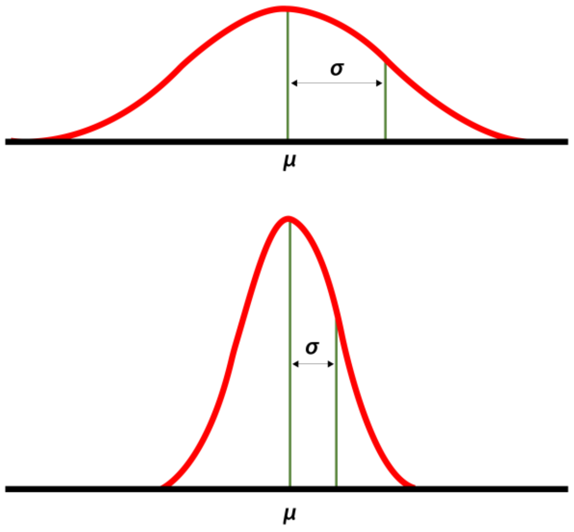
The signal-to-noise ratio can also be given as the ratio of the variance of the signal to the variance of the noise, or it can be written in any other manner that essentially emphasizes the level of variance between the signal and the noise.

Data variation

If one has the mean of the data and its variance or standard deviation, another meaningful quantity to represent data is the **coefficient of variation**:

$$V = \frac{standard\ deviation\ of\ data}{mean\ of\ data}$$

and can be expressed in percentage of dB.



Arithmetic mean of a data set $\{x_1, x_2, \ldots, x_N\}$:

$$\langle x \rangle = \mu = \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)$$

Standard deviation (or σ) is a measure of how dispersed the data is in relation to the mean. σ^2 is known as the variance.

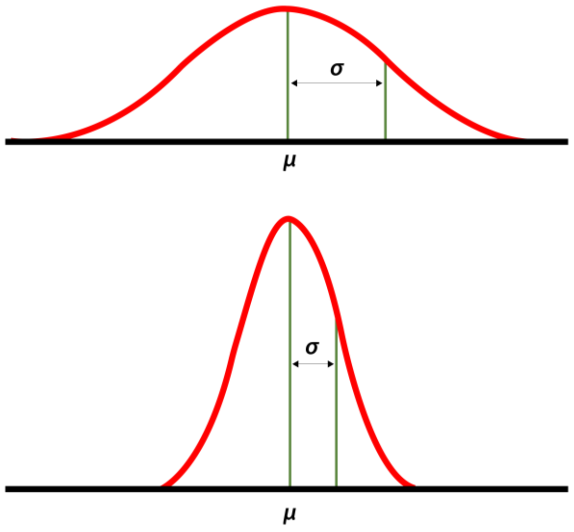
$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

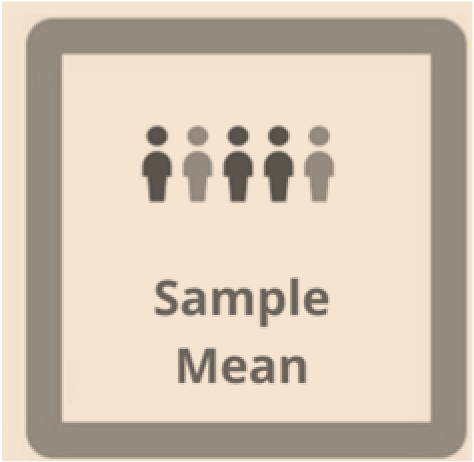
Data variation

If one has the mean of the data and its variance or standard deviation, another meaningful quantity to represent data is the **coefficient of variation**:

$$V = \frac{standard\ deviation\ of\ data}{mean\ of\ data}$$

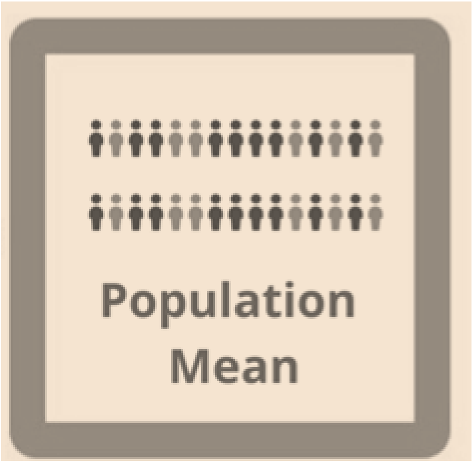
and can be expressed in percentage of dB.





$\bar{x}, s_x$

VERSUS



μ, σ

Practical considerations

Analysis means to separate something into components in order to identify, interpret and/or study the underlying structure of a problem. In order to do this, you should have some idea of what the components are likely to be (and be prepared to find new components!). Therefore, you should have some sort of model of the data in mind or at least past works to base some of your potential findings.

You need to have some idea as to what to look for in the data and how to analyze it. If you cannot figure out a technical basis for why you should use one method over another, then try several different methods and state how the results can be sensitive (or not) to distinct analysis methods. This can yield valuable insights regarding the data and/or analysis strategy to be adopted. Often, the analysis strategy will evolve based on the new insights gained during the discovery process.

Practical considerations

Note that models can be a kind of guide – it is important to keep an open mind about the data and analysis. Consider other interpretations, the potential for bias, data errors, and preconceived notions or assumptions influencing interpretations.

There is no data analysis (or data coding) without reading and studying. You will have to read (a lot) during your research being that experimental or computational in nature.

You should be your strongest critic when evaluating the quality of your analysis.

You have to know/understand which tools to apply for any particular analysis, and how to apply those tools. You also need to understand the result or outcome of the analysis and must make sure that the results are consistent with your <u>RESEARCH QUESTION</u> and <u>PROJECT GOALS</u>.

D. G. Martinson, Quantitative Methods of Data Analysis for the Physical Sciences and Engineering.