First Detailed Example: Data Layout

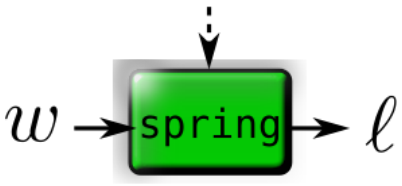
Ishan Paul

August 2022

In this section, we will learn the proper format to store the data, obtained from the experiment. Readers may be enticed to skip this section. However we must emphasize on the fact that no matter how good a model is, it cannot be analysed if the data is not stored in a way that is readable by standard software such as R, spss or sas. Therefore, this knowledge is indispensable.

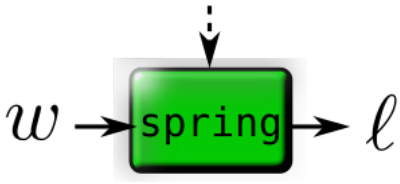
We use the example introduced in the previous section. Before proceeding, let's go over the set-up once again. We have a spring. Making sure that there is no external interference, we observe the length of the spring while it is being extended by a weight. We repeat this experiment n times. So, we have readings of lengths, $l_1, l_2, ..., l_n$ against weights, $w_1, w_2, ..., w_n$. All lengths must be measured in the same unit. Same goes for the weights. So, the data obtained from the experiment consists of n pairs, $(w_1, l_1), (w_2, l_2), ..., (w_n, l_n)$. Note that $w_1, ..., w_n$ need not be all distinct.

Now our work ***must*** begin by looking at the "black-box diagram".



Observe that apart from the random error, there are two arrows, for $\boldsymbol{w}$ as the input and $\boldsymbol{l}$ as the output. So, in this case, the data must be stored in a table made up of two columns, one for weights and one for lengths. We have taken 6 readings with weights, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 units. Here we have chosen equally spaced weights to suit our convenience. However, in a lab this is not necessary. In fact even repeating weights is not an uncommon practice.

W	L
1.0	5.29
1.5	6.31
2.0	7.28
2.5	8.33
3.0	9.30
3.5	10.32



In the above diagram, each column represents readings on the corresponding variable. In this case, we have only two columns for weight and length. In general, the number of columns is equal to the number of arrows in the "black-box diagram", excluding the random error, which is not observed. Each row represents a single observation of a weight and the corresponding length of the spring. Thus, number of rows in the table is equal to the number of observations taken in the experiment.

If the data is stored in a ".txt" file, it is advisable to store it directly with the headings in the first line as follows.

```
W     L
1.0   5.29
1.5   6.31
2.0   7.28
2.5   8.33
3.0   9.30
3.5   10.32
```

The data is then ready to be used with any software such as R, spss or sas. We conclude by reiterating the importance of the "black-box diagram" in designing the data layout. It greatly simplifies our work, especially in more complicated setups. So, we must look at the "black-box diagram", and then design the data layout accordingly.