

## 1.2: What is a Linear Regression Model?

Suppose that we have measured the performance of several different computer systems using some standard benchmark program. We can organize these measurements into a table, such as the example data shown in Table 1.1. The details of each system are recorded in a single row. Since we measured the performance of  $n$  different systems, we need  $n$  rows in the table.

Table 1.1: An example of computer system performance data.

System	Inputs			Output
	Clock (MHz)	Cache (kB)	Transistors (M)	Performance
1	1500	64	2	98
2	2000	128	2.5	134
...	...	...	...	...
i	...	...	...	...
n	1750	32	4.5	113

The first column in this table is the index number (or name) from 1 to  $n$  that we have arbitrarily assigned to each of the different systems measured. Columns 2-4 are the *input parameters*. These are called the *independent variables* for the system we will be modeling. The specific values of the

input parameters were set by the experimenter when the system was measured, or they were determined by the system configuration. In either case, we know what the values are and we want to measure the performance obtained for these input values. For example, in the first system, the processor's clock was 1500 MHz, the cache size was 64 kbytes, and the processor contained 2 million transistors. The last column is the performance that was measured for this system when it executed a standard benchmark program. We refer to this value as the *output* of the system. More technically, this is known as the system's *dependent variable* or the system's *response*.

The goal of regression modeling is to use these  $n$  independent measurements to determine a mathematical function,  $f()$ , that describes the relationship between the input parameters and the output, such as:

$$\text{performance} = f(\text{Clock}, \text{Cache}, \text{Transistors})$$

This function, which is just an ordinary mathematical equation, is the regression model. A regression model can take on any form. However, we will restrict ourselves to a function that is a linear combination of the input parameters. We will explain later that, while the function is a linear combination of the input parameters, the parameters themselves do not need to be linear. This linear combination is commonly used in regression modeling and is powerful enough to model most systems we are likely to encounter.

In the process of developing this model, we will discover how important each of these inputs are in determining the output value. For example, we might find that the performance is heavily dependent on the clock frequency, while the cache size and the number of transistors may be much less important. We may even find that some of the inputs have essentially no impact on the output making it completely unnecessary to include them in the model. We also will be able to use the model we develop to predict the performance we would expect to see on a system that has input values that did not exist in any of the systems that we actually measured. For instance, Table 1.2 shows three new systems that were not part of the set of systems that we previously measured. We can use our regression model to predict the performance of each of these three systems to replace the question marks in the table.

Table 1.2: An example in which we want to predict the performance of new systems  $n + 1$ ,  $n + 2$ , and  $n + 3$  using the previously measured results from the other  $n$  systems.

System	Inputs			Output
	Clock (MHz)	Cache (kB)	Transistors (M)	Performance
1	1500	64	2	98

2	2000	128	2.5	134
...	...	...	...	...
i	...	...	...	...
...	...	...	...	...
n	1750	32	4.5	113
n + 1	2500	256	2.8	?
n + 2	1560	128	1.8	?
n + 3	900	64	1.5	?

As a final point, note that, since the regression model is a linear combination of the input values, the values of the model parameters will automatically be scaled as we develop the model. As a result, the units used for the inputs and the output are arbitrary. In fact, we can rescale the values of the inputs and the output before we begin the modeling process and still produce a valid model.

---

This page titled [1.2: What is a Linear Regression Model?](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [David Lilja \(University of Minnesota Libraries Publishing\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.