Talking about distribution is a generic topic as we can refer to two different mathematical tools:

* Probability distribution
* Frequency distribution

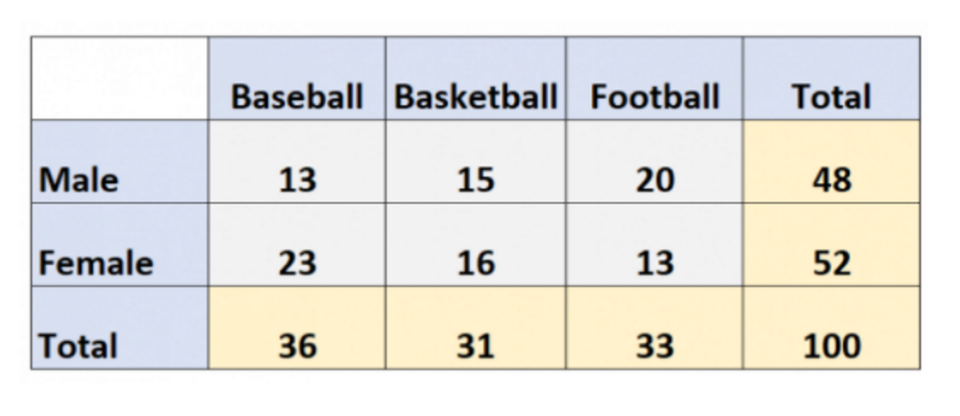
The probability distribution in [probability theory](https://en.wikipedia.org/wiki/Probability_theory)  is the mathematical [function](https://en.wikipedia.org/wiki/Function_(mathematics)) that gives the probabilities of occurrence of different possible outcomes for an [experiment](https://en.wikipedia.org/wiki/Experiment_(probability_theory)). It is a mathematical description of a [random](https://en.wikipedia.org/wiki/Randomness) phenomenon in terms of its [sample space](https://en.wikipedia.org/wiki/Sample_space) and the [probabilities](https://en.wikipedia.org/wiki/Probability) of [events](https://en.wikipedia.org/wiki/Event_(probability_theory)).

The second is typically used within a statistical context and is a representation, either in a graphical or tabular format, that displays the number of observations within a given interval. The frequency is how often a value occurs in an interval. Analysts often use a frequency distribution to visualize or illustrate the data collected in a sample. For example, the height of children can be split into several different categories or ranges. In measuring the height of 50 children, some are tall and some are short, but there is a high probability of a higher frequency or concentration in the middle range. The most important factors for gathering data are that the intervals used must not overlap and must contain all the possible observations. Frequency distributions can be presented as a frequency table, a histogram, or a bar chart.

<https://www.investopedia.com/terms/f/frequencydistribution.asp>

<https://en.wikipedia.org/wiki/Probability_distribution>

Before talking about the different types of distribution we must introduce the concept of **Bivariate Setting**. Unlike the *Univariate* one, here we have **two**attributes to work with (*X, Y*) and the results are simply the combination of the two attributes.



In the figure above we have the representation of this type of setting whose two variables *X, Y* are respectively the *Gender*and the *Sport Practiced*.

Even if we have more than one attribute, we can still consider them **separated** and calculate the *Univariate Distribution*of each. For instance, the univariate distribution of *Male*is *48*, while the one for *Baseball* is *36*. As you can see, the result is shown in the **last** **row** and **column** of the scheme, which is called **Marginal Distribution**. It represents the *sum* of the elements in that row or column (= the frequency of value assumed by attribute X).

In the same table, we can notice another type of distribution: **Conditional**. The conditional distribution, is the *set of values* for each row (or column) and it’s called *Conditional* because all these values have a class of attribute in common, therefore we are *conditioning* those values to **that** particular category. By looking at the picture, we can for example say that the set *13, 15, 20* is a conditional distribution of *Sport Practiced* given *X = Male*.

References:

<https://machinelearningmastery.com/joint-marginal-and-conditional-probability-for-machine-learning/>

<https://sites.nicholas.duke.edu/statsreview/jmc/>