**Introduction to Random Variables**

Previously we defined an ***experiment*** as any process that generates will defined outcomes.

We can describe the outcomes of an experiment using numbers and this numerical description is called a random variable.

|  |
| --- |
| A random variable is a numerical description of the outcome of the experiment. |

*Remark: We will call random variables R.V.s for short.*

We associate a number with each outcome of the experiment and the particular numerical value of the random variable depends on the outcome.

For example, consider the experiment of picking an employee at random from an office. We are interested in the number of years work experience the employee has.

The outcome of the experiment could be one year, two years and so on. The random variable is the number of experience the employee has and the value of the random variable can change every time we repeat the experiment.  
  
  
We can repeat the experiment of picking an employee at random many times and count the number of times that the outcome was one year, two years and so on.

Recall that the probability ***p = r/n*** where

* **n** is the number of time we carry out an experiment (i.e. “number of trials”)
* **r** is the number of times we get the result we are interested in (i.e. “number of successes”)

We can therefore associate probabilities with the various possible values that the random variable might take.

The values of the random variable and the associated probabilities form a probability distribution.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Years (R.V.)** | **1** | **2** | **3** | **4** | **5** |
| **Probability** | **0.2** | **0.3** | **0.3** | **0.1** | **0.1** |

The value of the random variable is unknown before we carry out the experiment but using the probability distribution in the table above, we can say that the probability that an employee selected at random will have one years experience is ***0.2*** or ***20%.***

Although experiments have outcomes that are naturally described using numbers, other do not. For example, an employee might be asked if they enjoy their work. The random variable is enjoyment of work, and the outcomes are yes and no.

We can arbitrarily assign the value ***zero*** to an employee who says no and ***one*** to an employee who says yes.

In this way, the random variable still provides a numerical description of the outcome of the experiment.

Many methods of statistical analysis assume that data collected follows a known probability distribution. We will later examine the most commonly used probability distributions.