Simpson's Diversity Index is used to calculate a measure of diversity, taking into account the number of something as well as its abundance. The index is most often used for ecological studies that measure species diversity, but the same analysis can also be applied to other principles, such as the diversity of opinion on an idea over a geographical space.

Why would we use Simpson's diversity index?

Researchers frequently need to know if the range of data they produce is showing a true level of variety or if it just appears to do so on paper. This is especially the case when researchers are dealing with very large quantities of data and the level of diversity within that data is not easy to decipher from reading a table of results.

The index measures the probability that two randomly selected individuals from a sample will be the same. The formula for calculating the value of the index (D) is

$$D = 1 - \frac{\Sigma n(n-1)}{N(N-1)}$$

where n is the number of individuals displaying one trait (e.g. the number of individuals of one species)
N = the total number of all individuals

Worked Example:

To best use this calculation, it is a good idea to tabulate the findings. In this example, the different species seen in quadrats at two sand dune sites are compared.

Species	Location A quadrat			Location B quadrat		
	n (number of individuals)	n-1	$n(n-1)$	n	n-1	$n(n-1)$
Sea couch grass	21	20	420	18	17	306
Sea twitch grass	1	0	0	11	10	110
Sea rocket	8	7	56	8	7	56
Prickly saltwort	1	0	0	3	2	6
Lyme grass	12	11	132	7	6	42
N (total number of individuals)	43			47		
		Σ	608		Σ	520

For Location A:

$D = 1 - \frac{608}{43 \times 42}$ $D = 1 - \frac{608}{1806}$ $D = 1 - 0.337$ $D = 0.663$

For Location B:

$D = 1 - \frac{520}{47 \times 46}$ $D = 1 - \frac{520}{2162}$ $D = 1 - 0.241$ $D = 0.759$

The value of *D* will always fall between 0 and 1, where 1 represents complete diversity and 0 represents complete uniformity.

One index value on its own holds very little value: but once the researcher is able to compare it to another, it can begin to show something. For example, in the above example, the higher value at Location B shows that from the data collected, Location B appears to be more diverse in species than Location A.