Frequency Distributions

Learning Objective: Summarize and describe the distribution of a categorical variable in context.

What is your perception of your own body? Do you feel that you are overweight, underweight, or about right?

A random sample of 1,200 U.S. college students were asked this question as part of a larger survey. The following table shows part of the responses:

Body Image

Student	Body Image
student 25	overweight
student 26	about right
student 27	underweight
student 28	about right
student 29	about right

Here is some information that would be interesting to get from these data:

- What percentage of the sampled students fall into each category?
- How are students divided across the three body image categories? Are they equally divided? If not, do the percentages follow some other kind of pattern?

There is no way that we can answer these questions by looking at the raw data, which are in the form of a long list of 1,200 responses, and thus not very useful. However, both these questions will be easily answered once we summarize and look at the **distribution** of the variable Body Image (i.e., once we summarize how often each of the categories occurs).

In order to summarize the distribution of a categorical variable, we first create a table of the different values (categories) the variable takes, how many times each value occurs (count) and, more importantly, how often each value occurs (by converting the counts to percentages); this table is called a frequency distribution. Here is the frequency distribution for our example:

Body Image Distribution

Category	Count	Percent
About right	855	$\left(\frac{855}{1200}\right) * 100 = 71.3\%$
Overweight	235	$\left(\frac{235}{1200}\right) * 100 = 19.6\%$
Underweight	110	$\left(\frac{110}{1200}\right) * 100 = 9.2\%$
Total	**n=1200**	**100%**