Illustration of the difference between correlation and interaction amongst independent variables[*]

Rajagopalan Ranganathan

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 3.0[†]

Abstract

Students often find it difficult to understand and articulate the difference between correlation and interaction between independent variables while building regression models. This illustration seeks to clarify these distinct aspects

In a particular modeling context, independent variables

- May or may not be correlated with each other
- May or may not have interactive effects on the dependent variable

One does not necessarily imply the other. However, students find it difficult to grasp this distinction. We clarify this distinction through a simple numerical example.

We are exploring how gender (male/female) and smoking status (smoker, non-smoker) influences the chances of cancer. That is, Gender and smoking status are our independent variables. The dependent variable is whether or not someone has cancer.

Correlation but no interaction

Consider the data in Table 1 below. The cell entries are the sample counts.

Table 1: Correlation			
	Smokers	Non-smokers	Total
Men	90	60	150
Women	60	90	150
Total	150	150	300

Table 1

Gender and smoking status are correlated, as proportionately more men are smokers, relative to women. We do not need any information on the dependent variable to say this.

Now look at the number cases of cancer in each category as given in Table 2 below.

Table 2: Cancer Counts			
	Smokers	Non-smokers	Total
Men	27	6	33
Women	36	18	54
Total	63	24	87

Table 2

Do gender and smoking status interact in influencing occurrence of cancer? To assess this clearly, let us convert the data in Table 2 into percentages of the corresponding cell entries in Table 1

Table 3: Cancer Rates		
	Smokers	Non-smokers
Men	30%	10%
Women	60%	20%

Table 3

Table 3 clearly says that smokers are thrice as likely as non-smokers to get cancer, irrespective of whether they are men or women. Similarly, women are twice as likely to get cancer as men, irrespective of whether they are smokers or non-smokers. In short, there is no 'interaction' between gender and smoking status as regards their impact on cancer rates.

No Correlation but Interaction

Now consider data Table 4 below. It is clear that there is no correlation between gender and smoking status. Both men and women have 50-50 chance being a smoker or non-smoker.

Table 4: No Correlation			
	Smokers	Non-smokers	Total
Men	90	90	180
Women	60	60	120
Total	150	150	300

Table 4

Let us assume the sample cancer counts as in Table 5 below.

Table 5: Cancer Counts			
	Smokers	Non-smokers	Total
Men	27	9	36
Women	36	24	60
Total	63	33	96

Table 5

Do gender and smoking status interact in influencing occurrence of cancer? To assess this clearly, let us convert the data in Table 2 into percentages of the corresponding cell entries in Table 4

Table 6: Cancer Rates		
	Smokers	Non-smokers
Men	30%	10%
Women	60%	40%

Table 6

Table 6indicates interaction. The impact of gender depends on smoking status. For example, cancer rates among women are twice that of men amongst smokers, but four times that of men amongst non-smokers. Similarly, smokers have thrice the cancer rates of non-smokers amongst men. But amongst women, this is only one and a half times.

In other words, the impact of one variable on cancer rates depends on the level of the other variable- a result of 'interaction'.