College Mathematics

two-way tables and conditional probability

A two-way table is simply a table that shows data classified in two different ways.
Here is an example of a two-way table, showing U.S. Senators classified by party affiliation and by gender.

United States Senators (113th Congress)

	men	women
Democrats	39	16
Republicans	41	4

How many of the Senators are men?
How many of the Senators are women?
How many of the Senators are Democrats?
How many of the Senators are Republicans?

What fraction of women in the Senate are Democrats?
What percent of women in the Senate are Democrats?

What fraction of Democrats in the Senate are women?
What percent of Democrats in the Senate are women?

Is it true that most women in the Senate are Democrats?
Is it true that most Democrats in the Senate are women?
Do you see that these are two *very* different questions?

Conditional probability is the probability of an event occurring, given knowledge of some other event.
The knowledge of the other event usually reduces the number of possible outcomes.

Here's an example of this idea, with a conclusion that is surprising to most people:

Assume that there is a rare medical condition that only occurs in 1 out of 1000 people.
Assume that the laboratory test that detects this medical condition is 99.8% accurate.
More precisely, out of 1000 test results, 998 will be correct; only 2 will be incorrect.

Suppose a person is randomly selected, and this test is performed.
Suppose the test result is positive (the test detected the condition).
What is the probability that the person actually *has* the condition?

Since the test is extremely accurate, it's natural to think that the probability is very high.

Actually, the probability depends on both the accuracy of the test and the rarity of the condition.
In this example, the probability is actually fairly low.

Here's one way to understand the situation more clearly:

Consider 1,000,000 randomly selected people.

Only 1000 of them have the medical condition,
while 999,000 of the people don't have the condition.

Of the 1000 people who have the condition, we would expect 998 to test positive (correctly) and
2 to test negative (incorrectly). The 2 incorrect negative results are called false negative results.

Of the 999,000 people who do not have the condition, we would expect 997,002 to test negative (correctly) and
1,998 to test positive (incorrectly). The 2 incorrect positive results are called false positive results.

Make a two-way table showing the possible cases.

	have the condition	don't have the condition
test positive	998	1,998
test negative	2	997,002
	1000	999,000

Given that the test was positive, the person must be in the top row.

The probability that the person has the condition, given that the test is positive, is $\frac{998}{998+1998}$, about $\frac{1}{3}$.

Is it true that most people *who have the condition* test positive?
Is it true that most people *who test positive* have the condition?

Do you see that these are two *very* different questions?