

## 4.7: Two-Way Tables (5 of 5)

### Learning Objectives

- Create a hypothetical two-way table to answer more complex probability questions.

In our previous work with probability, we computed probabilities using a two-way table of data from a large sample. Now we create a hypothetical two-way table to answer more complex probability questions.

### Example

#### Will It Be a Boy or a Girl?

A pregnant woman often opts to have an ultrasound to predict the gender of her baby.

Assume the following facts are known:

- Fact 1: 48% of the babies born are female.
- Fact 2: The proportion of girls correctly identified is 9 out of 10.
- Fact 3: The proportion of boys correctly identified is 3 out of 4.

(Source: Keeler, Carolyn, and Steinhorst, Kirk. "New Approaches to Learning Probability in the First Statistics Course," *Journal of Statistics Education* 9(3):1–24, 2001.)

Here are the questions we want to answer:

**Question 1:** *If the examination predicts a girl, how likely is it that the baby will be a girl?*

**Question 2:** *If the examination predicts a boy, how likely is it that the baby will be a boy?*

Let's consider what the possibilities are.

- The ultrasound examination predicts a girl, and either (a) a girl is born or (b) a boy is born.
- The ultrasound exam predicts a boy, and either (a) a girl is born or (b) a boy is born.

Let's represent these four possible outcomes in a two-way table. On the left we have the categorical variable *prediction*, and on the top the categorical variable *gender of baby*.

	Girl	Boy	
Predict Girl			
Predict Boy			

Now we find ourselves in an interesting situation. A two-way table without data!

The key idea is to create a two-way table consistent with the stated facts, then use the table to answer our questions.

To get started, let's assume we have ultrasound predictions for 1,000 random babies. We could have picked any number here, but 1,000 will make our calculations easier to keep track of.

Starting with this number, we work backwards with our three facts to fill in this "hypothetical" table.

The first step is to put 1,000 as the overall total in the bottom right corner.

	Girl	Boy	Row Totals
Predict Girl			
Predict Boy			
Column Totals			1,000

Let's consider Fact 1: 48% of the babies born are female.

The bottom row gives the distribution of the categorical variable *gender of baby*. We can use this fact to compute the total number of girls and boys.

- 48% girls means that  $0.48(1,000) = 480$  are girls.
- 52% are boys. (If 48% are girls, then  $100\% - 48\% = 52\%$  are boys.) So,  $0.52(1,000) = 520$  boys.

Fill these values into the bottom row of table.

- Note: These are marginal totals.
- You can check your work: These numbers should add to 1,000. If we add all the girls and boys together, we get the total number of babies.

	Girl	Boy	Row Totals
<b>Predict Girl</b>			
<b>Predict Boy</b>			
<b>Column Totals</b>	$0.48(1,000) = 480$	$0.52(1,000) = 520$	1,000

Now let's move on to Fact 2: The proportion of girls correctly identified is 9 out of 10.

- 9 out of 10 is 90% ( $9 \div 10 = 0.90 = 90\%$ ).
- 90% of the girls are correctly identified:  $0.90(480) = 432$ .
- 10% of the girls are misidentified (predicted to be a boy):  $0.10(480) = 48$ .

Fill these values into the table.

- You can check your work: These numbers should add to the total number of girls.
- (Girls who are correctly identified as girls) + (Girls who are misidentified as boys) = Total girls

	Girl	Boy	Row Totals
<b>Predict Girl</b>	$0.90(480) = 432$		
<b>Predict Boy</b>	$0.10(480) = 48$		
<b>Column Totals</b>	480	520	1,000

Finally, we use Fact 3: The proportion of boys correctly identified is 3 out of 4.

- 3 out of 4 is 75% ( $3 \div 4 = 0.75 = 75\%$ ).
- 75% of the boys are correctly identified:  $0.75(520) = 390$ .
- 25% of the boys are misidentified (predicted to be a girl):  $0.25(520) = 130$ .

Fill these values into the table.

- You can check your work: These numbers should add to the total number of boys.
- (Boys who are correctly identified as boys) + (Boys who are misidentified as girls) = Total boys

	Girl	Boy	Row Totals
<b>Predict Girl</b>	432	$0.25(520) = 130$	
<b>Predict Boy</b>	48	$0.75(520) = 390$	
<b>Column Totals</b>	480	520	1,000

Filling in the Row Totals, we now have a complete hypothetical two-way table based on our given information.

	Girl	Boy	Row Totals
<b>Predict Girl</b>	432	130	562
<b>Predict Boy</b>	48	390	438
<b>Column Totals</b>	480	520	1,000

Predict Boy	48	390	438
Column Totals	480	520	1,000

We are now in a position to answer our two questions:

**Question 1:** *If the examination predicts a girl, how likely is it that the baby will be a girl?*

**Answer:** We are asked to find the probability of a girl **given that** the examination predicts a girl.

This is the conditional probability:  $P(\text{girl} \mid \text{predict girl})$ .

So our answer to Question 1 is  $P(\text{girl} \mid \text{predict girl}) = 432 / 562 = 0.769$ .

**Question 2:** *If the examination predicts a boy, how likely is it that the baby will be a boy?*

**Answer:** We are asked to find the probability of a boy **given that** the examination predicts a boy.

This is the conditional probability:  $P(\text{boy} \mid \text{predict boy})$ .

So our answer to Question 2 is  $P(\text{boy} \mid \text{predict boy}) = 390 / 438 = 0.890$ .

**Conclusion:** If an ultrasound examination predicts a girl, the prediction is correct about 77% of the time. In contrast, when the prediction is a boy, it is correct 89% of the time.

#### Comment

Are you surprised at the answers to these questions? Looking just at the three given facts, you might have intuitively expected a different result. This is exactly why a two-way table is so useful. It helps us organize the relevant information in a way that permits us to carry out a logical analysis. When it comes to probability, sometimes our intuition needs some help.

Use the following context for the next Try It activity.

A large company has instituted a mandatory employee drug screening program. Assume that the drug test used is known to be 99% accurate. That is, if an employee is a drug user, the test will come back positive (“drug detected”) 99% of the time. If an employee is a non-drug user, then the test will come back negative (“no drug detected”) 99% of the time. Assume that 2% of the employees of the company are drug users.

In constructing the hypothetical two-way table, it is convenient to start by assuming that the company has 10,000 employees (10,000 is a large enough number to ensure that all calculations result in whole numbers).

#### Try It

<https://assessments.lumenlearning.co...sessments/3545>

<https://assessments.lumenlearning.co...sessments/3546>

<https://assessments.lumenlearning.co...sessments/3547>

<https://assessments.lumenlearning.co...sessments/3548>

<https://assessments.lumenlearning.co...sessments/3549>

<https://assessments.lumenlearning.co...sessments/3550>

#### Contributors and Attributions

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by:** Open Learning Initiative. **Located at:** <http://oli.cmu.edu>. **License:** CC BY: Attribution

4.7: Two-Way Tables (5 of 5) is shared under a [not declared](#) license and was authored, remixed, and/or curated by LibreTexts.

- 5.7: Two-Way Tables (5 of 5) by Lumen Learning is licensed CC BY 4.0.