

5.6: Two-Way Tables (4 of 5)

Learning Objectives

- Analyze and compare risks using conditional probabilities.

When we calculate the probability of a **negative outcome** like a heart attack, we often refer to the probability as a **risk**. For example, we talk about the *probability* of winning the lottery but the *risk* of getting struck by lightning. Whenever you see the word *risk*, keep in mind it's just another word for *probability*.

Example

Risk and the Physicians' Health Study

Researchers in the Physicians' Health Study (1989) designed a randomized clinical trial to determine whether aspirin reduces the risk of heart attack. Researchers randomly assigned a large sample of healthy male physicians (22,071) to one of two groups. One group took a low dose of aspirin (325 mg every other day). The other group took a placebo. This was a double-blind experiment. Here are the final results.

	Heart Attack	No Heart Attack	Row Totals
Aspirin	139	10,898	11,037
Placebo	239	10,795	11,034
Column Totals	378	21,693	22,071

Note that the categorical variables in this case are

- Explanatory variable*: Treatment (aspirin or placebo)
- Response variable*: Medical outcome (heart attack or no heart attack)

Question: Does aspirin lower the risk of having a heart attack?

To answer this question, we compare two conditional probabilities:

- The probability of a heart attack given that aspirin was taken every other day.
- The probability of a heart attack given that a placebo was taken every other day.

From the table we have

- $P(\text{heart attack} \mid \text{aspirin}) = 139 / 11,037 = 0.013$
- $P(\text{heart attack} \mid \text{placebo}) = 239 / 11,034 = 0.022$

The result shows that taking aspirin reduced the risk from 0.022 to 0.013.

We often compare two risks by calculating the **percentage change**. We calculate the difference (how much the risk changed) and divide by the risk for the placebo group.

Here is the calculation:

$$\frac{\text{frac}\{\text{0.013}-\text{0.022}\}\{\text{0.022}\}}{\text{frac}\{\text{0.009}\}\{\text{0.022}\}}\approx\text{0.41}$$

Therefore, we conclude that taking aspirin results in a 41% reduction in risk.

As reported in the *New England Journal of Medicine*, "This trial of aspirin for the primary prevention of cardiovascular disease demonstrates a conclusive reduction in the risk of myocardial infarction (heart attack)." (Source: "Final Report on the Aspirin Component of the Ongoing Physicians' Health Study," *New England Journal of Medicine* 321(3):129–35, 1989.)

Comment

In the preceding example, we compared the difference in risk (how much the risk changed) to the risk for the placebo (nontreatment) group:

$$\text{percentage reduction of risk} = \frac{\text{new treatment risk} - \text{placebo risk}}{\text{placebo risk}}$$

In general, we are interested in determining how much a new treatment reduces the risk compared to a **reference** risk. The reference may be nontreatment (e.g., use of a placebo), or it could be an existing treatment that we hope to improve on. So we have:

$$\text{percentage reduction of risk} = \frac{\text{new treatment risk} - \text{reference risk}}{\text{reference risk}}$$

The following table is used for the next Learn By Doing activity.

	Nonfatal	Fatal	Row Totals
Seat Belt	412,368	510	412,878
No Seat Belt	162,527	1,601	164,128
Column Totals	574,895	2,111	577,006

Learn By Doing

<https://assessments.lumenlearning.co...sessments/3542>

<https://assessments.lumenlearning.co...sessments/3543>

<https://assessments.lumenlearning.co...sessments/3544>

Let's summarize our work with probability. We defined three kinds of probabilities related to a two-way table.

- A *marginal probability* is the probability of a categorical variable taking on a particular value *without regard to the other categorical variable*. For example, $P(\text{Health Sciences})$ is the probability that a student is enrolled in the Health Sciences program. In calculating the probability, we use overall student data contained in the margins of the table. We do not take into account the other categorical variable: gender.
- A *conditional probability* is the probability of a categorical variable taking on a particular value *given the condition that the other categorical variable has some particular value*. For example, $P(\text{Health Sciences given female})$ is the probability that a student is enrolled in Health Sciences given that we know the student is female. In calculating the probability, we use only a subset of the data. The subset used is determined by the given condition: if our condition relates to female students, then we consider only the information in the table pertaining to females.
- A *joint probability* is the probability that the *two categorical variables each take on a specific value*. For example: $P(\text{male and Info Tech})$ is the probability that a student is both a male and in the Info Tech program. In calculating this probability, we divide the count in one inner cell of the table by the overall total count (in the lower right corner).

When we calculate the probability of a negative outcome like a heart attack, we often refer to the probability as a *risk*. We compare risk by calculating the percentage change:

$$\text{percentage reduction of risk} = \frac{\text{new treatment risk} - \text{reference risk}}{\text{reference risk}}$$

CC licensed content, Shared previously

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

This page titled [5.6: Two-Way Tables \(4 of 5\)](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Bill Pelz](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.