Establishing Independence with Data

Data often serve as a basis for establishing a probability model or for checking whether an assumed model is reasonable. In using data to check for independence, however, you have to be careful. You might ask, "Why do you have to be careful?" In this activity you'll answer that question.

COLLECT DATA

1. Record whether you are right-handed or left-handed.
2. Determine (and record) if you are right-eyed or left-eyed. Hold your hands together in front of you at arm's length. Make a space between your hands that you can see through. Through the space, look at an object at least 15 feet away. Now close your right eye. Can you still see the object? If so, you are left-eyed. Now close your left eye. Can you still see the object? If so, you are right-eyed.

Q1 Would you expect being right-handed and being right-eyed to be independent? That is, if you know a person is right-handed, does that change the probability that he or she is right-eyed?

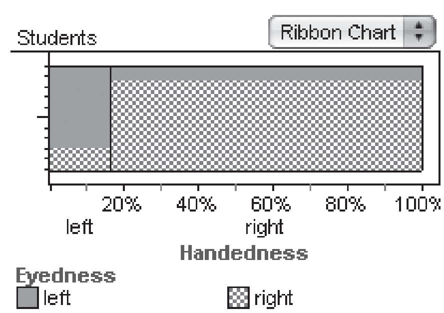
INVESTIGATE

Hands and Eyes

3. Enter the data for your class into a Fathom document. Make a two-way table (that is, a summary table) with *Handedness* along one table dimension and *Eyedness* along the other.

Click on a section of the ribbon chart. (Hold down the Shift key to click additional sections.) Move your cursor over the *collection.* You can read the number of selected cases in the lower-left corner of the Fathom window.

4. Make a bar chart for *Handedness.* Change it into a ribbon chart, then drag *Eyedness* into the *interior* of the graph. (*Eyedness* is now the legend.) The entire bar represents your whole class. Each slice represents a different handedness. The two sections of each slice represent left-eyed and right-eyed.



Q2 What area on your ribbon chart represents the proportion $P(right\text{-}eyed \mid right\text{-}handed)$?

Q3 Based on your ribbon chart, compare these two proportions: $P(\textit{right-eyed} \mid \textit{right-handed})$ and $P(\textit{right-eyed})$. Which proportion is larger, or are they equal?

Q4 After examining the two-way table and the ribbon chart, would you say that being right-handed and being right-eyed are independent?

Q5 Use the definition of independence to check your answer to Q4.

Coin Flips

Now you'll simulate two coin flips and study the results.

Consider randomInteger or randomPick as functions for generating coin flips.

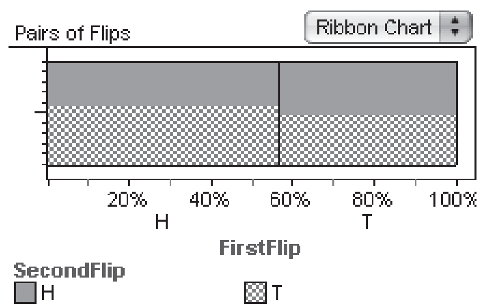
5. Create a collection with 100 cases and two attributes, *FirstFlip* and *SecondFlip*. Use formulas to randomly generate heads or tails for each attribute.

Pairs of Flips

	FirstFlip	SecondFlip
1	T	H
2	H	T
3	T	T
4	T	H
5	T	H

Q6 Would you expect the results of the first flip and the second flip to be independent? That is, if you know the outcome of the first flip, does that change the probabilities for the second flip?

6. Make a two-way table and a ribbon chart to analyze your data.



Q7 What area on your ribbon chart represents the proportion $P(\textit{SecondFlip is heads} \mid \textit{FirstFlip is heads})$?

Q8 Based on your ribbon chart, compare these two proportions: $P(\textit{SecondFlip is heads} \mid \textit{FirstFlip is heads})$ and $P(\textit{SecondFlip is heads})$. Which proportion is larger, or are they equal?

Q9 After examining the two-way table and the ribbon chart, would you say that the results of the first flip and the second flip are independent?

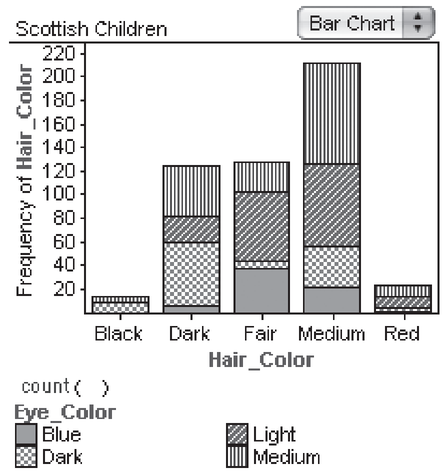
Q10 Use the definition of independence to check your answer to Q9.

Q11 Did you get the results you expected in Q10? Explain.

Q12 What does a ribbon chart look like when two attributes are independent? When they are not independent?

EXPLORE MORE

1. Change your ribbon chart to a bar chart (stacked bar graph). What does a stacked bar graph look like when two attributes are independent? When they are not independent?



2. Open the Fathom document **Scottish Children.ftm.** Make a ribbon chart or stacked bar graph and determine if the attributes *Hair_Color* and *Eye_Color* are independent. Explain your reasoning.
3. Make a simulation of eyedness and handedness in which the attributes aren't independent.
4. Invent a measure of how the results in a two-way table differ from what you would expect if the two attributes are independent. It might be, for example, that if you get exactly what you expect, the value of the measure is zero and the measure increases the farther the results are from what you expect. Apply your measure to your collections from the activity.
5. Using your measure from Explore More 4, repeatedly collect measures for your coin-flip collection and build a distribution of measures for many sets of 100 flips.

Establishing Independence with Data — Activity Notes

Objectives

- Recognizing the difficulty of establishing independence with real-world data
- Using the definition of independent events, $P(A \mid B) = P(A)$, and the multiplication rule for independent events, $P(A \text{ and } B) = P(A) \cdot P(B)$, to check for independence
- Understanding that because the two proportions in any experiment are probably not equal, $P(A \mid B) \neq P(A)$ does not necessarily mean that the events are not independent
- Seeing that data do not often show independence clearly but sometimes data can demonstrate that independence is not a reasonable assumption

Activity Time: 30–50 minutes (25 minutes as a presentation)

Setting: Paired/Individual Activity (collect data or use **Independence.ftm**) or Whole-Class Presentation (use **IndependencePresent.ftm**)

Optional Document: ScottishChildren.ftm (Explore More 2)

Statistics Prerequisites

- Definition of probability
- Familiarity with the definition of independent events, $P(A \mid B) = P(A)$
- Familiarity with the concept of independent events
- Familiarity with the conditional $P(A \mid B)$

Statistics Skills

- Multiplication rule
- Working with the definition of conditional probability
- Working with the definition of independence
- Comparing probabilities with graphs and two-way tables
- Recognizing independence or dependence with graphs and two-way tables
- *Optional:* Stacked bar charts (Explore More 1 and 2)

AP Course Topic Outline: Part III A (1, 3, 5), B (1)

Fathom Prerequisites: Students should be able to make case tables, summary tables, and ribbon charts and define attributes based on a random function.

Fathom Skills: Students use summary tables to numerically check for independence, use ribbon charts to visually check for independence, use a summary table as a two-way table, work with ribbon charts to compare probabilities, and work with legends in ribbon charts.

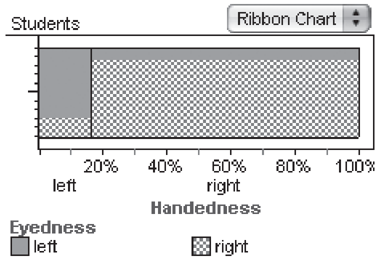
General Notes: This activity demonstrates that samples often do not meet the mathematical definition of independence even when we have reason to believe that the events are independent. Sometimes, however, data can demonstrate that independence is not a reasonable assumption.

Procedure: Eye dominance and hand dominance are associated, so a random sample should produce data that look dependent: $P(\textit{right-eyed} \mid \textit{right-handed}) > P(\textit{right-eyed})$. One note of caution: The percentage of left-handers is rather small, so you may need to have a fairly large group of students to see any left-handed students.

The coin flips are independent, and the data should support this even though students are not likely to find that $P(H \text{ and } H)$ is exactly equal to $P(H \textit{ on first flip}) \cdot P(H \textit{ on second flip})$ in any set of sample data. If you don't have time for your students to collect the data, you can have them use the Fathom document **Independence.ftm** instead. It has a collection of data for 30 students that your students can use to work through the activity. Alternatively, **IndependencePresent.ftm** has everything made. With this document projected, you could do the questions as a class activity.

INVESTIGATE

Q2 The large checkered area on this ribbon chart represents the proportion $P(\textit{right-eyed} \mid \textit{right-handed})$. Go to the section that is right-handed, then look at just the right-eyed group in that section.

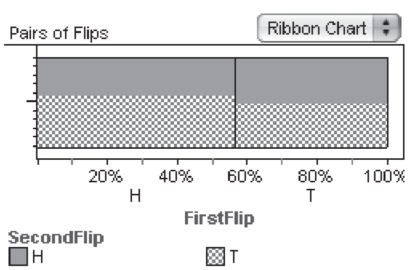


Q3 The proportion of the shaded region above that represents $P(\textit{right-eyed} \mid \textit{right-handed})$ compared to the whole section of right-handedness is larger than the proportion of *right-eyed* to the whole (both sections). (You might want to remind students that they are looking at proportions, not at counts). Using the table: There are 21 right-eyed students in the right-handed row. That proportion is 21/25. The proportion for right-eyed is $22/30 \approx 0.73$.

Q4–Q5 The table and ribbon chart suggest dependence, not independence. Using the definition: $P(\textit{right-eyed} \mid \textit{right-handed}) = 21/25 = 0.84 > 22/30 = P(\textit{right-eyed})$. Using the multiplication rule: $P(\textit{right-eyed} \text{ and } \textit{right-handed}) = 21/30 = 0.7$, but $P(\textit{right-eyed}) \cdot P(\textit{right-handed}) = 22/30 \cdot 25/30 \approx 0.6$.

Q7 The top-right area below is the proportion $P(\textit{SecondFlip is heads} \mid \textit{FirstFlip is heads})$.

Q8 Based on the ribbon chart below, the two proportions $P(\textit{SecondFlip is heads} \mid \textit{FirstFlip is heads})$ and $P(\textit{SecondFlip is heads})$ look like they could be equal. They aren't exactly equal—which is the problem.



Pairs of Flips

		SecondFlip H	SecondFlip T	Row Summary
FirstFlip	H	28	26	54
	T	28	18	46
Column Summary		56	44	100

S1 = count( )

Q9–Q10 The coin flips are independent but they will fail the test. Using the definition: $P(\textit{SecondFlip is heads} \mid \textit{FirstFlip is heads}) = 28/54 \approx 0.52$ and $P(\textit{SecondFlip}) \approx 56/100 = 0.56$. Using the multiplication rule: $P(\textit{SecondFlip is heads} \text{ and } \textit{FirstFlip is heads}) = 28/100 = 0.28$, but $P(\textit{SecondFlip is heads}) \cdot P(\textit{FirstFlip is heads}) = 56/100 \cdot 54/100 = 0.3024$. Not exactly equal but close.

Q12 A ribbon chart will have roughly the same *proportion* of each category in each section when the attributes are independent.

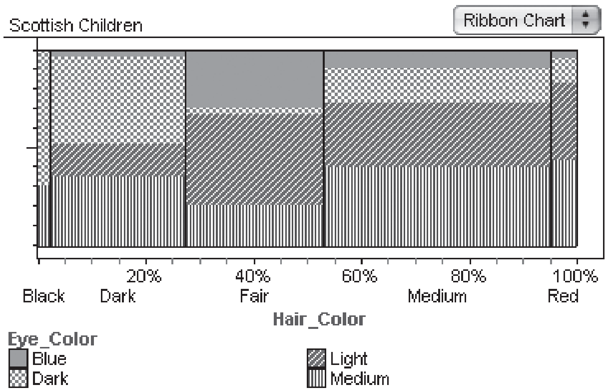
DISCUSSION QUESTIONS

- What does a ribbon chart look like when two attributes are independent? When they are not independent?
- What did you find about the independence of the first flip and the second flip? How much variation from "expected" did you find?

EXPLORE MORE

1. A stacked bar chart will have roughly the same *proportion* of each category in each section when the attributes are independent. This is harder to see in a stacked bar chart because the bar chart works with counts and the ribbon chart works with proportions.
2. There is an association between hair color and eye color, so they are not independent. As can be seen in the ribbon chart, the proportions of the legend attribute are not the same across all sections. For example, the fair-haired children have a greater proportion of light eyes than do, say, the dark-haired

children. So, just looking at the light-eyed rectangle in each section, it is clearly proportionally not the same across the different hair colors.



4. This question emphasizes the need for a statistical test of independence. Chi-square, χ^2, is one such measure. For now, leave this as an open-ended extension and allow students to create their own measures—there is no right or wrong answer.