3.5: Contingency Tables

A *contingency table* provides a way of portraying data that can facilitate calculating probabilities. The table helps in determining conditional probabilities quite easily. The table displays sample values in relation to two different variables that may be dependent or contingent on one another. Later on, we will use contingency tables again, but in another manner.

Example 3.5.1

Suppose a study of speeding violations and drivers who use cell phones produced the following fictional data:

	Speeding violation in the last year	No speeding violation in the last year	Total
Cell phone user	25	280	305
Not a cell phone user	45	405	450
Total	70	685	755

The total number of people in the sample is 755. The row totals are 305 and 450. The column totals are 70 and 685. Notice that 305 + 450 = 755 and 70 + 685 = 755.

Calculate the following probabilities using the table.

a. Find $P(\text{Person is a car phone user})$.
b. Find $P(\text{person had no violation in the last year})$.
c. Find $P(\text{Person had no violation in the last year AND was a car phone user})$.
d. Find $P(\text{Person is a car phone user OR person had no violation in the last year})$.
e. Find $P(\text{Person is a car phone user GIVEN person had a violation in the last year})$.
f. Find $P(\text{Person had no violation last year GIVEN person was not a car phone user})$

Answer

a. $\dfrac{\text{number of car phone users}}{\text{total number in study}} = \dfrac{305}{755}$
b. $\dfrac{\text{number that had no violation}}{\text{total number in study}} = \dfrac{685}{755}$
c. $\dfrac{280}{755}$
d. $\left(\dfrac{305}{755} + \dfrac{685}{755}\right) - \dfrac{280}{755} = \dfrac{710}{755}$
e. $\dfrac{25}{70}$ (The sample space is reduced to the number of persons who had a violation.)
f. $\dfrac{405}{450}$ (The sample space is reduced to the number of persons who were not car phone users.)

Exercise 3.5.1

Table shows the number of athletes who stretch before exercising and how many had injuries within the past year.

	Injury in last year	No injury in last year	Total
Stretches	55	295	350
Does not stretch	231	219	450
Total	286	514	800

a. What is $P(\text{athlete stretches before exercising})$?
b. What is $P(\text{athlete stretches before exercising} \mid \text{no injury in the last year})$?

Answer

a. $P(\text{athlete stretches before exercising}) = \frac{350}{800} = 0.4375$

b. $P(\text{athlete stretches before exercising} | \text{no injury in the last year}) = \frac{295}{514} = 0.5739$

Example 3.5.2

Table shows a random sample of 100 hikers and the areas of hiking they prefer.

Hiking Area Preference

Sex	**The Coastline**	**Near Lakes and Streams**	**On Mountain Peaks**	**Total**
Female	18	16	___	45
Male	___	___	14	55
Total	___	41	___	___

a. Complete the table.

b. Are the events "being female" and "preferring the coastline" independent events? Let *F* = being female and let *C* = preferring the coastline.

1. Find *P*(*F* AND *C*).
2. Find *P*(*F*)*P*(*C*)
3. Are these two numbers the same? If they are, then *F* and *C* are independent. If they are not, then *F* and *C* are not independent.

c. Find the probability that a person is male given that the person prefers hiking near lakes and streams. Let $\text{M} =$ being male, and let $\text{L} =$ prefers hiking near lakes and streams.

1. What word tells you this is a conditional?
2. Fill in the blanks and calculate the probability: *P*(___|___) = ___.
3. Is the sample space for this problem all 100 hikers? If not, what is it?

d. Find the probability that a person is female or prefers hiking on mountain peaks. Let $\text{F} =$ being female, and let $\text{P} =$ prefers mountain peaks.

1. Find $P(\text{F})$.
2. Find $P(\text{P})$.
3. Find $P(\text{F AND P})$.
4. Find $P(\text{F OR P})$.

Answers

a.

Hiking Area Preference

Sex	The Coastline	Near Lakes and Streams	On Mountain Peaks	Total
Female	18	16	**11**	45
Male	**16**	**25**	14	55
Total	**34**	41	**25**	**100**

b.

$$P(\text{F AND C}) = \frac{18}{100} = 0.18$$

$$P(\text{F})P(\text{C}) = \left(\frac{45}{100}\right)\left(\frac{34}{100}\right) = (0.45)(0.34) = 0.153$$

$P(\text{F AND C}) \neq P(\text{F})P(\text{C})$, so the events F and C are not independent.

c.

1. The word 'given' tells you that this is a conditional.
2. $P(\text{M}|\text{L}) = \frac{25}{41}$
3. No, the sample space for this problem is the 41 hikers who prefer lakes and streams.

d.

a. Find $P(\text{F})$.
b. Find $P(\text{P})$.
c. Find $P(\text{F AND P})$.
d. Find $P(\text{F OR P})$.

d.

1. $P(\text{F}) = \frac{45}{100}$
2. $P(\text{P}) = \frac{25}{100}$
3. $P(\text{F AND P}) = \frac{11}{100}$
4. $P(\text{F OR P}) = \frac{45}{100} + \frac{25}{100} - \frac{11}{100} = \frac{59}{100}$

Exercise 3.5.2

Table shows a random sample of 200 cyclists and the routes they prefer. Let $\text{M} =$ males and $\text{H} =$ hilly path.

Gender	Lake Path	Hilly Path	Wooded Path	Total
Female	45	38	27	110
Male	26	52	12	90
Total	71	90	39	200

a. Out of the males, what is the probability that the cyclist prefers a hilly path?
b. Are the events “being male” and “preferring the hilly path” independent events?

Answer

a. $P(H|M) = \frac{52}{90} = 0.5778$
b. For M and H to be independent, show $P(H|M) = P(H)$
$P(H|M) = 0.5778$, $P(H) = \frac{90}{200} = 0.45$
$P(H|M)$ does not equal $P(H)$ so M and H are NOT independent.

Example 3.5.3

Muddy Mouse lives in a cage with three doors. If Muddy goes out the first door, the probability that he gets caught by Alissa the cat is $\frac{1}{5}$ and the probability he is not caught is $\frac{4}{5}$. If he goes out the second door, the probability he gets caught by Alissa is $\frac{1}{4}$ and the probability he is not caught is $\frac{3}{4}$. The probability that Alissa catches Muddy coming out of the third door is $\frac{1}{2}$ and the probability she does not catch Muddy is $\frac{1}{2}$. It is equally likely that Muddy will choose any of the three doors so the probability of choosing each door is $\frac{1}{3}$.

Door Choice

Caught or Not	Door One	Door Two	Door Three	Total

Caught or Not	Door One	Door Two	Door Three	Total
Caught	$\frac{1}{15}$	$\frac{1}{12}$	$\frac{1}{6}$	____
Not Caught	$\frac{4}{15}$	$\frac{3}{12}$	$\frac{1}{6}$	____
Total	____	____	____	1

- The first entry $\frac{1}{15} = \left(\frac{1}{5}\right)\left(\frac{1}{3}\right)$ is $P(\text{Door One AND Caught})$
- The entry $\frac{4}{15} = \left(\frac{4}{5}\right)\left(\frac{1}{3}\right)$ is $P(\text{Door One AND Not Caught})$

Verify the remaining entries.

a. Complete the probability contingency table. Calculate the entries for the totals. Verify that the lower-right corner entry is 1.
b. What is the probability that Alissa does not catch Muddy?
c. What is the probability that Muddy chooses Door One OR Door Two given that Muddy is caught by Alissa?

Solution

Door Choice

Caught or Not	Door One	Door Two	Door Three	Total
Caught	$\frac{1}{15}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{19}{60}$
Not Caught	$\frac{4}{15}$	$\frac{3}{12}$	$\frac{1}{6}$	$\frac{41}{60}$
Total	$\frac{5}{15}$	$\frac{4}{12}$	$\frac{2}{6}$	1

b. $\frac{41}{60}$
c. $\frac{9}{19}$

Example 3.5.4

Table contains the number of crimes per 100,000 inhabitants from 2008 to 2011 in the U.S.

United States Crime Index Rates Per 100,000 Inhabitants 2008–2011

Year	Robbery	Burglary	Rape	Vehicle	Total
2008	145.7	732.1	29.7	314.7	
2009	133.1	717.7	29.1	259.2	
2010	119.3	701	27.7	239.1	
2011	113.7	702.2	26.8	229.6	
Total					

TOTAL each column and each row. Total data = 4,520.7

a. Find $P(\text{2009 AND Robbery})$.
b. Find $P(\text{2010 AND Burglary})$.
c. Find $P(\text{2010 OR Burglary})$.
d. Find $P(\text{2011}|\text{Rape})$.

e. Find $P(\text{Vehicle}|2008)$

Answer

a. 0.0294, b. 0.1551, c. 0.7165, d. 0.2365, e. 0.2575

Exercise 3.5.3

Table relates the weights and heights of a group of individuals participating in an observational study.

Weight/Height	Tall	Medium	Short	Totals
Obese	18	28	14	
Normal	20	51	28	
Underweight	12	25	9	
Totals				

a. Find the total for each row and column
b. Find the probability that a randomly chosen individual from this group is Tall.
c. Find the probability that a randomly chosen individual from this group is Obese and Tall.
d. Find the probability that a randomly chosen individual from this group is Tall given that the idividual is Obese.
e. Find the probability that a randomly chosen individual from this group is Obese given that the individual is Tall.
f. Find the probability a randomly chosen individual from this group is Tall and Underweight.
g. Are the events Obese and Tall independent?

Answer

Weight/Height	Tall	Medium	Short	Totals
Obese	18	28	14	60
Normal	20	51	28	99
Underweight	12	25	9	46
Totals	50	104	51	205

a. Row Totals: 60, 99, 46. Column totals: 50, 104, 51.
b. $P(\text{Tall}) = \frac{50}{205} = 0.244$
c. $P(\text{Obese AND Tall}) = \frac{18}{205} = 0.088$
d. $P(\text{Tall}|\text{Obese}) = \frac{18}{60} = 0.3$
e. $P(\text{Obese}|\text{Tall}) = \frac{18}{50} = 0.36$
f. $P(\text{Tall AND Underweight}) = \frac{12}{205} = 0.0585$
g. No. $P(\text{Tall})$ does not equal $P(\text{Tall}|\text{Obese})$.

References

1. “Blood Types.” American Red Cross, 2013. Available online at http://www.redcrossblood.org/learn-a...od/blood-types (accessed May 3, 2013).
2. Data from the National Center for Health Statistics, part of the United States Department of Health and Human Services.
3. Data from United States Senate. Available online at www.senate.gov (accessed May 2, 2013).
4. Haiman, Christopher A., Daniel O. Stram, Lynn R. Wilkens, Malcom C. Pike, Laurence N. Kolonel, Brien E. Henderson, and Loīc Le Marchand. “Ethnic and Racial Differences in the Smoking-Related Risk of Lung Cancer.” The New England Journal of

Medicine, 2013. Available online at http://www.nejm.org/doi/full/10.1056/NEJMoa033250 (accessed May 2, 2013).
5. “Human Blood Types.” Unite Blood Services, 2011. Available online at http://www.unitedbloodservices.org/learnMore.aspx (accessed May 2, 2013).
6. Samuel, T. M. “Strange Facts about RH Negative Blood.” eHow Health, 2013. Available online at http://www.ehow.com/facts_5552003_st...ive-blood.html (accessed May 2, 2013).
7. “United States: Uniform Crime Report – State Statistics from 1960–2011.” The Disaster Center. Available online at http://www.disastercenter.com/crime/ (accessed May 2, 2013).

Chapter Review

There are several tools you can use to help organize and sort data when calculating probabilities. Contingency tables help display data and are particularly useful when calculating probabilities that have multiple dependent variables.

Use the following information to answer the next four exercises. Table shows a random sample of musicians and how they learned to play their instruments.

Gender	Self-taught	Studied in School	Private Instruction	Total
Female	12	38	22	72
Male	19	24	15	58
Total	31	62	37	130

Exercise 3.5.4

Find P(musician is a female).

Exercise 3.5.5

Find $P(\text{musician is a male AND had private instruction})$.

Answer

$$P(\text{musician is a male AND had private instruction}) = \frac{15}{130} = \frac{3}{26} = 0.12$$

Exercise 3.5.6

Find P(musician is a female OR is self taught).

Exercise 3.5.7

Are the events “being a female musician” and “learning music in school” mutually exclusive events?

Answer

$$P(\text{being a female musician AND learning music in school}) = \frac{38}{130} = \frac{19}{65} = 0.29$$

$$P(\text{being a female musician})\text{P}(\text{learning music in school}) = \left(\frac{72}{130}\right)\left(\frac{62}{130}\right) = \frac{4,464}{16,900} = \frac{1,116}{4,225} = 0.26$$

No, they are not independent because $P(\text{being a female musician AND learning music in school})$ is not equal to $P(\text{being a female musician})P(\text{learning music in school})$.

Bringing it Together

Use the following information to answer the next seven exercises. An article in the *New England Journal of Medicine*, reported about a study of smokers in California and Hawaii. In one part of the report, the self-reported ethnicity and smoking levels per day were given. Of the people smoking at most ten cigarettes per day, there were 9,886 African Americans, 2,745 Native Hawaiians, 12,831 Latinos, 8,378 Japanese Americans, and 7,650 Whites. Of the people smoking 11 to 20 cigarettes per day, there were 6,514 African Americans, 3,062 Native Hawaiians, 4,932 Latinos, 10,680 Japanese Americans, and 9,877 Whites. Of the people smoking 21 to 30 cigarettes per day, there were 1,671 African Americans, 1,419 Native Hawaiians, 1,406 Latinos, 4,715 Japanese

Americans, and 6,062 Whites. Of the people smoking at least 31 cigarettes per day, there were 759 African Americans, 788 Native Hawaiians, 800 Latinos, 2,305 Japanese Americans, and 3,970 Whites.

Exercise 3.5.8

Complete the table using the data provided. Suppose that one person from the study is randomly selected. Find the probability that person smoked 11 to 20 cigarettes per day.

Smoking Levels by Ethnicity

Smoking Level	African American	Native Hawaiian	Latino	Japanese Americans	White	TOTALS
1–10						
11–20						
21–30						
31+						
TOTALS						

Exercise 3.5.9

Suppose that one person from the study is randomly selected. Find the probability that person smoked 11 to 20 cigarettes per day.

Answer

$$\frac{35,065}{100,450}$$

Exercise 3.5.10

Find the probability that the person was Latino.

Exercise 3.5.11

In words, explain what it means to pick one person from the study who is "Japanese American **AND** smokes 21 to 30 cigarettes per day." Also, find the probability.

Answer

To pick one person from the study who is Japanese American AND smokes 21 to 30 cigarettes per day means that the person has to meet both criteria: both Japanese American and smokes 21 to 30 cigarettes. The sample space should include everyone in the study. The probability is $\frac{4,715}{100,450}$.

Exercise 3.5.12

In words, explain what it means to pick one person from the study who is "Japanese American **OR** smokes 21 to 30 cigarettes per day." Also, find the probability.

Exercise 3.5.13

In words, explain what it means to pick one person from the study who is "Japanese American **GIVEN** that person smokes 21 to 30 cigarettes per day." Also, find the probability.

Answer

To pick one person from the study who is Japanese American given that person smokes 21-30 cigarettes per day, means that the person must fulfill both criteria and the sample space is reduced to those who smoke 21-30 cigarettes per day. The probability is $\frac{4,715}{15,273}$.

Exercise 3.5.14

Prove that smoking level/day and ethnicity are dependent events.

Glossary

contingency table

the method of displaying a frequency distribution as a table with rows and columns to show how two variables may be dependent (contingent) upon each other; the table provides an easy way to calculate conditional probabilities.