4.6: Joint and Marginal Probabilities and 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 4.6.1

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

Answer

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

Example 4.6.2

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 $P(F \text{ AND } C)$.
   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 $\mathrm{M}=$ being male, and let $\mathrm{L}=$ 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 $\mathrm{F}=$ being female, and let $\mathrm{P}=$ prefers mountain peaks.
   1. Find $P(\mathrm{F})$.
   2. Find $P(\mathrm{P})$.
   3. Find $P(\mathrm{F} \text{ AND } \mathrm{P})$.
   4. Find $P(\mathrm{F} \text{ OR } \mathrm{P})$.

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.

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

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

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

c.

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

d.

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

d.

1. $P(\mathrm{F})=\frac{45}{100}$

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

Example 4.6.3

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	$\frac{1}{15}$	$\frac{1}{12}$	$\frac{1}{6}$	____
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 4.6.4

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 $P(2011|\text{Rape})$
e. Find $P(\text{Vehicle}|2008)$

Answer

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

References

1. “Blood Types.” American Red Cross, 2013. Available online at 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 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 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).

WeBWork Problems

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.