Chapter 3 - Probability

Donald Butler

09/16/2021

```
library(tidyverse)
library()
```

Dice rolls. (3.6, p. 92) If you roll a pair of fair dice, what is the probability of

(a) getting a sum of 1?

There are no ways to roll a 1 from a pair of dice, 0%.

(b) getting a sum of 5?

There are 4 ways to roll a 5, (1,4), (2,3), (3,2), (4,1) out of 36 possible outcomes.

```
4/36
```

```
## [1] 0.1111111
```

(c) getting a sum of 12?

There is 1 way to roll a 12, (6,6) out of 36 possible outcomes.

```
1/36
```

```
## [1] 0.02777778
```

Poverty and language. (3.8, p. 93) The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English (foreign language) at home, and 4.2% fall into both categories.

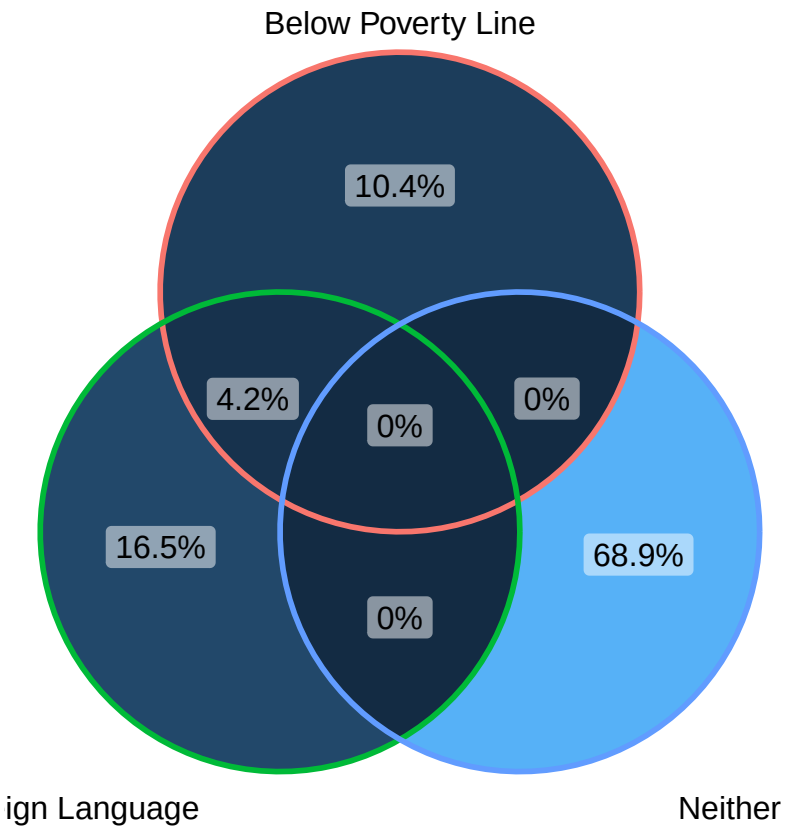
(a) Are living below the poverty line and speaking a foreign language at home disjoint?

No, 4.2% are both living below the poverty line and speaking a foreign language at home.

(b) Draw a Venn diagram summarizing the variables and their associated probabilities.

```
library(ggVennDiagram)

x <- list(BP = 1:146, FL = 105:311, EAP = 312:1000)
ggVennDiagram(x, category.names = c('Below Poverty Line','Foreign Language','Neither'),
              label_percent_digit = 1, label = 'percent') + theme(legend.position = 'none')
```



(c) What percent of Americans live below the poverty line and only speak English at home?

10.4%

(d) What percent of Americans live below the poverty line or speak a foreign language at home?

31.1%

(e) What percent of Americans live above the poverty line and only speak English at home?

68.9%

(f) Is the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home?

No, 25.3% of Americans living below the poverty line speak a foreign language at home. Similarly, 20.3% of Americans speaking a foreign language at home are living below the poverty line. There is some association between these two groups.

Assortative mating. (3.18, p. 111) Assortative mating is a nonrandom mating pattern where individuals with similar genotypes and/or phenotypes mate with one another more frequently than what would be expected under a random mating pattern. Researchers studying this topic collected data on eye colors of 204 Scandinavian men and their female partners. The table below summarizes the results. For simplicity, we only include heterosexual relationships in this exercise.

		Partner (female)			
		Blue	Brown	Green	Total
Self (male)	Blue	78	23	13	114
	Brown	19	23	12	54
	Green	11	9	16	36
	Total	108	55	41	204

(a) What is the probability that a randomly chosen male respondent or his partner has blue eyes?

```
(108 + 114 - 78)/204
```

```
## [1] 0.7058824
```

(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?

```
78/144
```

```
## [1] 0.5416667
```

(c) What is the probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes?

```
19/54
```

```
## [1] 0.3518519
```

What about the probability of a randomly chosen male respondent with green eyes having a partner with blue eyes?

```
11/36
```

```
## [1] 0.3055556
```

(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

The probability that a couple has the same color eyes is:

```
(78 + 23 + 16) / 204
```

```
## [1] 0.5735294
```

With 57% of couples choosing a partner with the same eye color, it does appear that they are not an independent choice.

Books on a bookshelf. (3.26, p. 114) The table below shows the distribution of books on a bookcase based on whether they are nonfiction or fiction and hardcover or paperback.

		Format		
		Hardcover	Paperback	Total
Type	Fiction	13	59	72
	Nonfiction	15	8	23
	Total	28	67	95

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.

```
28/95 * 59/94
```

```
## [1] 0.1849944
```

(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.

```
72/95 * 28/94
```

```
## [1] 0.2257559
```

(c) Calculate the probability of the scenario in part (b), except this time complete the calculations under the scenario where the first book is placed back on the bookcase before randomly drawing the second book.

```
72/95 * 28/95
```

```
## [1] 0.2233795
```

(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.

The probability of (c) is slightly less than (b) because of the chance of drawing the same book twice in (c).

Baggage fees. (3.34, p. 124) An airline charges the following baggage fees: $25 for the first bag and $35 for the second. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage and 12% have two pieces. We suppose a negligible portion of people check more than two bags.

(a) Build a probability model, compute the average revenue per passenger, and compute the corresponding standard deviation.

```
baggage_fees <- c(0,25,35)
baggage_prob <- c(.54,.34,.12)
(baggage_ev <- sum(baggage_fees*baggage_prob))
```

```
## [1] 12.7
```

```
(baggage_sd <- sum((baggage_fees - baggage_ev)^2)^.5)
```

```
## [1] 28.45821
```

(b) About how much revenue should the airline expect for a flight of 120 passengers? With what standard deviation? Note any assumptions you make and if you think they are justified.

```
(120 * baggage_ev)
```

```
## [1] 1524
```

Since the baggage fee cannot be less than zero and the standard deviation is more than double than the mean, it may suggest that the expected revenue from a flight be lower than the expected value of 1524.

This assumes that the number of bags brought by each passenger is independent of the other passengers. This may not be the case since people often fly in groups.

Income and gender. (3.38, p. 128) The relative frequency table below displays the distribution of annual total personal income (in 2009 inflation-adjusted dollars) for a representative sample of 96,420,486 Americans. These data come from the American Community Survey for 2005-2009. This sample is comprised of 59% males and 41% females.

Income	*Total*
\$1 to \$9,999 or loss	2.2%
\$10,000 to \$14,999	4.7%
\$15,000 to \$24,999	15.8%
\$25,000 to \$34,999	18.3%
\$35,000 to \$49,999	21.2%
\$50,000 to \$64,999	13.9%
\$65,000 to \$74,999	5.8%
\$75,000 to \$99,999	8.4%
\$100,000 or more	9.7%

(a) Describe the distribution of total personal income.

The distribution of total personal income is right skewed. The final bracket of \$100,000 or more contains significant outliers with very high incomes.

(b) What is the probability that a randomly chosen US resident makes less than \$50,000 per year?

```
.022 + .047 + .158 + .183 + .212
```

```
## [1] 0.622
```

(c) What is the probability that a randomly chosen US resident makes less than \$50,000 per year and is female? Note any assumptions you make.

```
.622 * .41
```

```
## [1] 0.25502
```

This assumes that the gender of the individual is independent of their personal income.

(d) The same data source indicates that 71.8% of females make less than \$50,000 per year. Use this value to determine whether or not the assumption you made in part (c) is valid.

This would indicate that there is an association between gender and personal income, so the assumption made in (c) is not valid.