Chapter 3 - Probability

Shana Green

DATA 606 - Homework 3

Due Date: 9/13/2020

```r
library(VennDiagram)
```

```
## Loading required package: grid
```

```
## Loading required package: futile.logger
```

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?
(b) getting a sum of 5?
(c) getting a sum of 12?

Since a fair dice has six sides, the total sample space is 6 x 6 = 36

a) Each fair dice has a digit from one to six, the probability of getting a sum of 1 is 0, since the lowest sum is two.

b) Sample space for the sum of 5: {(1,4),(2.3),(3,2),(4,1)}. The probability of getting a sum of 5 is 4/36.

c) Sample space for the sum of 12: {(6,6)}. The probability of getting a sum of 12 is 1/36.

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.

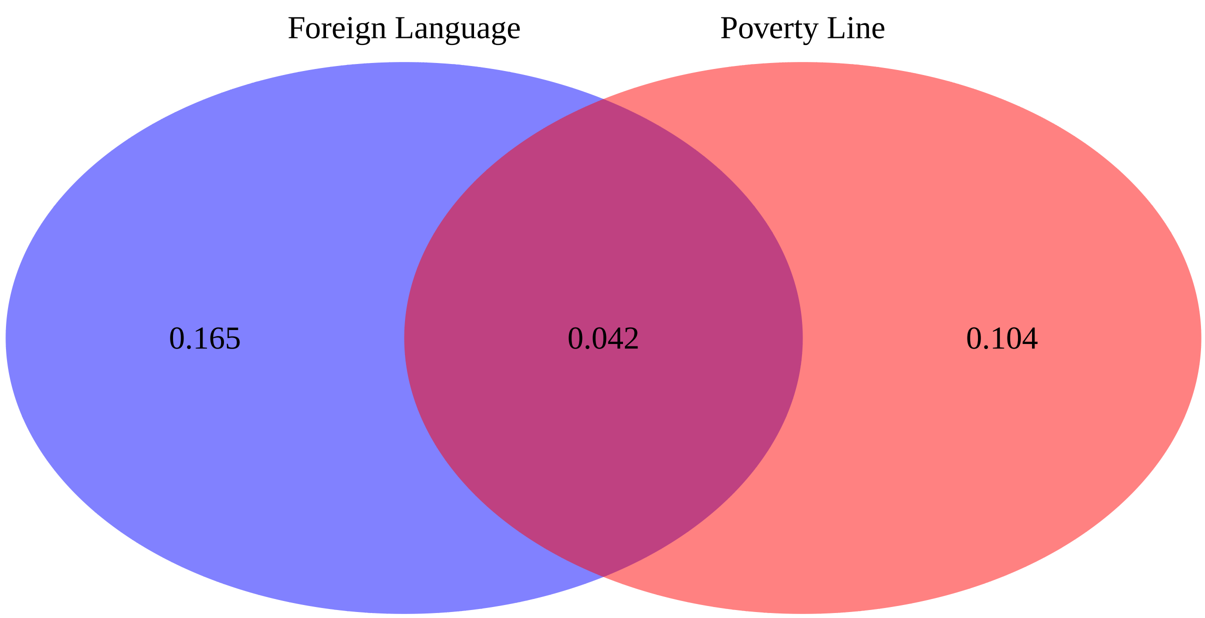
(a) Are living below the poverty line and speaking a foreign language at home disjoint? **No, there are people who are both living below the poverty line and speak a language other than English at home.**

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

```
# Foreign Language: (20.7 - 4.2)/100 = 0.165
# Poverty Line: (14.6 - 4.2)/100 = 0.104

library(grid)
grid.newpage()

draw.pairwise.venn(0.146, 0.207, 0.042, category = c("Poverty Line", "Foreign Language"), lty = "blank"
```




```
## (polygon[GRID.polygon.1], polygon[GRID.polygon.2], polygon[GRID.polygon.3], polygon[GRID.polygon.4],
```

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

Each person living below the poverty line either speak only English at home or does not. Since .146 of Americans live below the poverty line and .042 speak a language other than English at home, the other .104x100 = 10.4% only speak English at home.

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

P(below PL or speak FL)=P(below PL)+P(speak FL)-P(both)=0.146+0.207-0.042=0.311x100=31.1%

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

P(neither below PL nor speak FL)=1-P(below PL or speak FL)=1-0.311=0.689x100 =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?

P(below PL) x P(speak FL)=0.146x0.207=0.030x100=3%, which does not equal P(below PL and speak FL)=0.042x100=4.2%, therefore the events are dependent.

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?

114/204+108/204-(78/204)=144/204=0.71. The probability that either one of the partners to have blue eyes is 144/204.

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

78/114 = 0.68. The probability that the chosen female has blue eyes provided that the male partner has blue eyes is 78/114.

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

19/54 = 0.35. The probability that the chosen female has blue eyes provided that the male partner has brown eyes is 19/54.

11/36 = 0.31. The probability that the chosen female has blue eyes provided that the male partner has green eyes is 11/36.

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

78/114 is not equal to P(A)=10/204, hence the events in this problem are not independent.

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 x 59/94 = 0.185. The probability of drawing a hardcover book first, then a paperback fiction book is 0.185.

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

72/95 x 27/94 = 0.2177. The probability of drawing a fiction book first and then a hardcover book is 0.2177.

(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) x (28/95) = 0.2234

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

The events are similar and the sample size is large. Although the experiment is complete with replacement, and we drew only two books, the sample space is not affected greatly.

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.

```
# Number of bags
luggage <- c(0, 1, 2)

# Fee charged for 0, 1st, and 2nd luggage
bag0 <- 0
bag1 <- 25
bag2 <-bag1  + 35

# Fees table created
fees <- c(bag0,bag1,bag2)

# Percentage of passengers that check baggage in decimal form
bpb <- c(0.54, 0.34, 0.12)

# Find Expected value for each x_i
revenue <- fees * bpb

# mu
mu <- sum(revenue)

# Expected Revenue per passenger
mu
```

```
## [1] 15.7
```

```
# sd

bagvar <- fees - mu

bagvar2 <- bagvar^2 * bpb
varsquared <- sum(bagvar2)

# sd by calculating the square root of the variance
sd <- varsquared^(1/2)

sd
```

```
## [1] 19.95019
```

(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 passengers
revenue2 <- fees * bpb * 120
```

```r
# 2nd mu
mu2 <- sum(revenue2)

mu2
```

```
## [1] 1884
```

```r
bagvariance <- fees - mu2
bagvar3 <- bagvariance^2 * bpb
varsquared2 <-sum(bagvar3)

sd2 <-varsquared2^(1/2)
sd2
```

```
## [1] 1868.407
```

The expected revenue for the 120 passengers is $1884.00. The standard deviation will be $1868.41.

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 is right skewed with a median between \$35,000 and \$49,999.

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

P(less than \$50,000) = 2.2 + 4.7 + 15.8 + 18.3 + 21.2 = 0.622 = 62.2%

(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.

Assuming that gender and income are independent: 0.622 x 0.41 = 0.255 = 25.5%

(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.

0.718 is not equal to 0.255, which concludes these events are not independent.