Chapter 3 - Probability

Md Forhad Akbar

Loading packages

```
library(tidyverse)
library(VennDiagram)
library(grid)
```

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?
Answer: probability of getting a sum of 1 from a pair of fair dice is 0 since there is no possible way to get a sum of 1 from a pair of fair dice. The least sum possible from a pair of fair dice is 2.

(b) getting a sum of 5?
Answer: probability of getting a sum of 5 from a pair of fair dice is 4/36= 11.11%. A sum of 5 can be possible in 4 ways, 1+4, 2+3, 3+2, 4+1, and the total possible outcomes are 36.

(c) getting a sum of 12?
Answer: probability of getting a sum of 5 from a pair of fair dice is 1/36= 2.78%. A sum of 12 can be possible in only one way, 6+6, and the total possible outcomes are 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?
Answer: No, since there are 4.2% of the population living below the poverty line and speaking a language other than English.

(b) Draw a Venn diagram summarizing the variables and their associated probabilities.
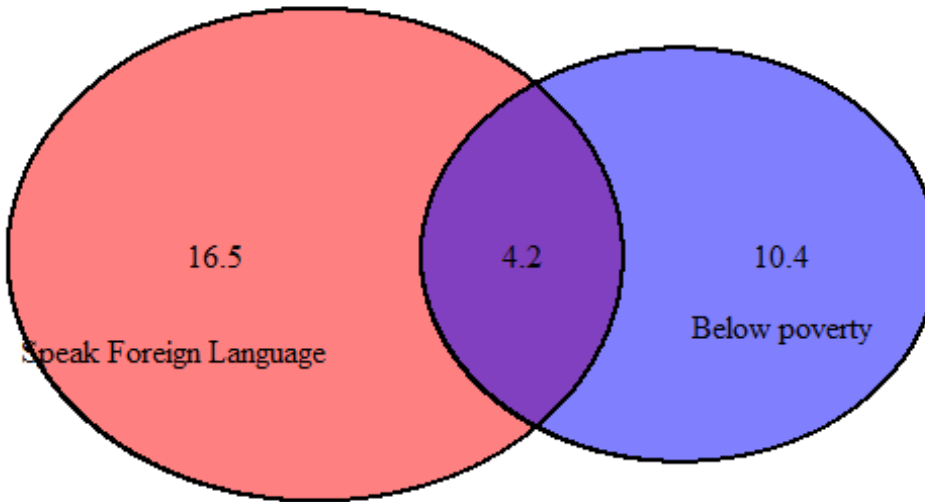
```
below_poverty <- 14.6
foreign_language <- 20.7
Both <- 4.2
poverty_only <- below_poverty - Both
foreign_language_only <- foreign_language - Both

plot <- draw.pairwise.venn(below_poverty,
                           foreign_language,
```

```
                              cross.area=Both,
                              c("Below poverty", "Speak Foreign Language"),
                              fill=c("blue", "red"),
                              cat.dist=-0.08,
                              ind=FALSE)
grid.draw(plot)
```


(c) What percent of Americans live below the poverty line and only speak English at home?
     Answer: Percentage of Americans live below the proverty line and only spreak English at home is 14.6% - 4.2% = 10.4%.

(d) What percent of Americans live below the poverty line or speak a foreign language at home?
     Answer: percentage of Americans live below the poverty line or speak a foreign language at home 14.6% + 20.7% - 4.2% = 31.1%

(e) What percent of Americans live above the poverty line and only speak English at home?
     Answer: Percentage of Americans live above the poverty line and only speak English at home 100% - 31.1% = 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?
     Answer: No, Those events are not independent, there is many possibilities that one variable falls in the other one.

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.

(a) What is the probability that a randomly chosen male respondent or his partner has blue eyes?
Answer: Lets assume
Male with Blue eyes = P(MB)
Partner Female with Blue eyes= P(FB)
Both Male Female with Blue eyes = P(MB AND FB)
male respondent or his partner has blue = P(MB or FB)
According to addition rule
P(MB or FB)=P(MB) + P(FB) - P(MB AND FB) = 114/204+108/204-78/204= 70.59%

(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?
Answer: The probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes is 78/114= 68.42%

(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?
Answer: Probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes is 19/54 = 35.19%
Probability of a randomly chosen male respondent with green eyes having a partner with blue eyes is 11/36 = 30.56%

(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.
Answer: The probability that a random self male from the study has a blue eyes is 56% and the probability of randomly select a partner female with blue eyes is 52.4% . Both variables are dependent.

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.

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.
Answer: P(Hardcover first) x P(paperback fiction) = (28/95) x (59 / 94) = 18.50%

(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.
Anwer: probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement is 72/95 x 28/94 = 22.58%

(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.
Answer:P(first: fiction and second: hardcover) = (72/95) x (28/95) = 22.34%

(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.
Answer: Because the difference of (b) and (c) is replacement of the book which changes the denominator from 1/(9594) to (1/9595). The impact of the overall percentage number is small.

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.

```
fees <- c(0,25,60)
passengers <- c(54,34,12)
baggage <- data.frame(fees,passengers)
# Average revenue
average_revenue <-
(sum((baggage$fees*baggage$passengers)))/sum(baggage$passengers)
average_revenue

## [1] 15.7

# Standard deviation
sqrt((0-average_revenue)^2*.54+(25-average_revenue)^2*.34+(60-average_revenue
)^2*.12)

## [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.

```
#Total airline expected revenue for 120 passengers

(120*.34)*25 + (120*.12)*35 +(120*.12)*25

## [1] 1884
```

```r
#Standard deviation
sqrt((0 -15.7)^2*0.54*120+(25 -15.7)^2*0.34*120+(60 -15.7)^2*0.12*120)

## [1] 218.5434
```

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.

(a) Describe the distribution of total personal income.

```r
income <- c("$1 to $9,999","$10,000 to $14,999","$15,000 to $24,999","$25,000
to $34,999","$35,000 to $49,999","$50,000 to $64,999","$65,000 to
$74,999","$75,000 to $99,999","$100,000 or more")

total <- c(2.2,4.7,15.8,18.3,21.2,13.9,5.8,8.4,9.7)

frequency_table <- data.frame(income,total)
frequency_table

##                 income total
## 1        $1 to $9,999   2.2
## 2 $10,000 to $14,999   4.7
## 3 $15,000 to $24,999  15.8
## 4 $25,000 to $34,999  18.3
## 5 $35,000 to $49,999  21.2
## 6 $50,000 to $64,999  13.9
## 7 $65,000 to $74,999   5.8
## 8 $75,000 to $99,999   8.4
## 9    $100,000 or more   9.7

barplot(frequency_table$total)
```
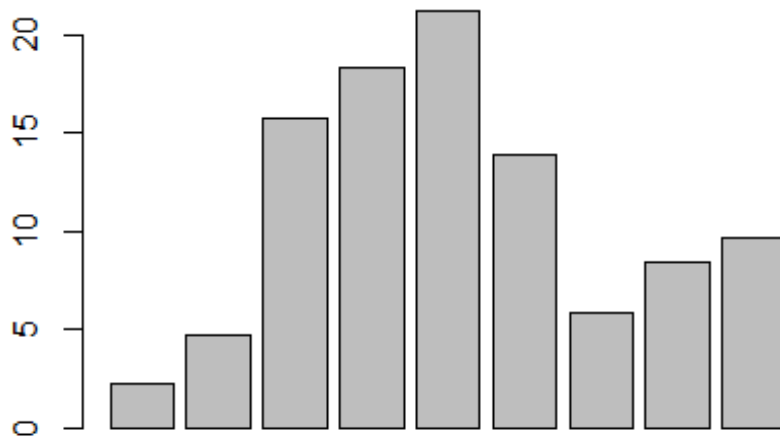
This distribution has dual peak and approximately normal. The mean is centered around 25K to 49K.

(b) What is the probability that a randomly chosen US resident makes less than $50,000 per year?
Answer: Probability that a randomly chosen US resident makes less than $50,000 per year is 21.2% + 18.3% + 15.8% + 4.7% + 2.2 = 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.
Answer: Probability that a randomly chosen US resident makes less than $50,000 per year and is femaleis 62.2 x 0.41 = 25.5%. We are assuming female population is evenly distributed.

(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.
Answer: If the source indicates that 71.8% of females make less than $50,000 per year then the assumption we make that female population is evenly distributed in not correct.