Chapter 3 - Probability

Subhalaxmi Rout

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?

** Answer **

(a) Probability of getting a sum of 1: 0 (beacuse two dice can give atleast sum of 2)
(b) Probability of getting a sum of 5: 4/36 or 1/9 [(1,4),(2,3),(3,2),(4,1)]
(c) Probability of getting a sum of 12: 1/36 [(6,6)]

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?
(b) Draw a Venn diagram summarizing the variables and their associated probabilities.
(c) What percent of Americans live below the poverty line and only speak English at home?
(d) What percent of Americans live below the poverty line or speak a foreign language at home?
(e) What percent of Americans live above the poverty line and only speak English at home?
(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 **

(a) No because 4.2% fall in to both categories.

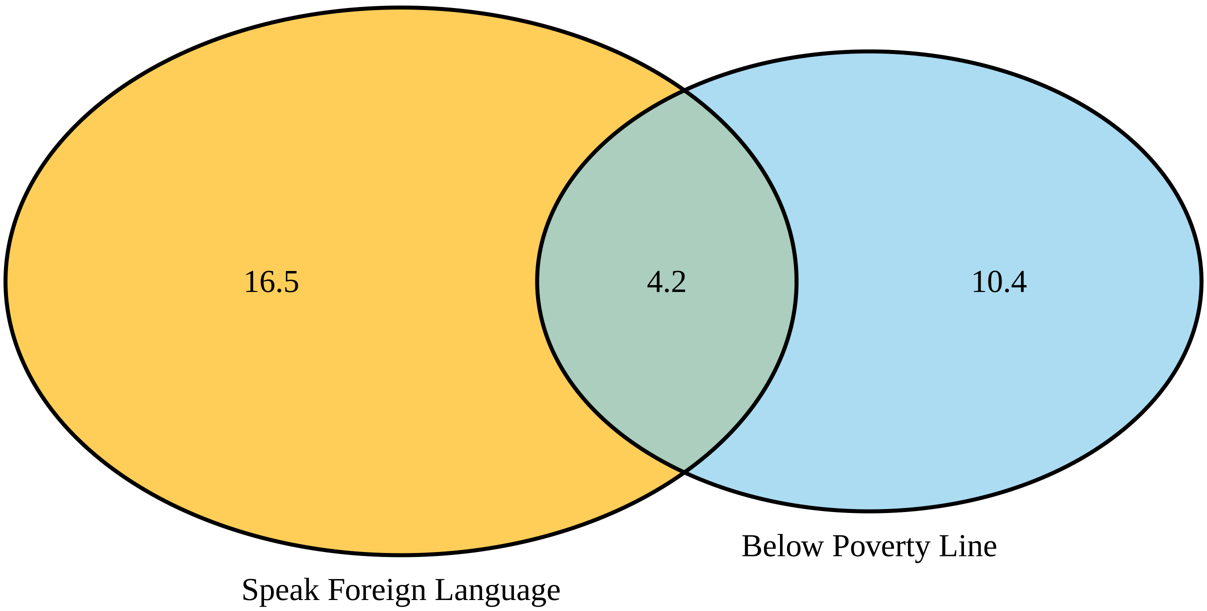
(b)

```r
#install.packages('VennDiagram')
library("VennDiagram")
```

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

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

```r
library("grid")
library("futile.logger")
grid.newpage()
draw.pairwise.venn(14.6, 20.7, 4.2, category = c("Below Poverty Line", "Speak Foreign Language"), fill =
```




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

```
(c) Below Poverty = 14.6
    Speak Only English = 14.6 - 4.2 = 10.4
    10.4 percent of Americans live below the poverty line and only speak English at home
(d) Below Poverty = 14.6
    Speak Foreign Language = 20.7
    P(A or B) = P(A) + P(B) - P(A and B)
              = 14.6 + 20.7 - 4.2
              = 31.1
    31.1 percent of Americans live below the poverty line or speak a foreign language at home
(e) Below poverty = 14.6
    Foreign Language speaker = 20.7
    Both(poverty and forign language speaker) = 4.2
    Above poverty and only speak english  = 100 - (16.5 + 4.2 + 10.4) = 100 - (31.1)
         = 68.9
    68.9 percent of Americans live above the poverty line and only speak English at home

(f) Yes, the events are independent.
```

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?
(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?
(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?
(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

** Answer **

```
(a) P(A) = 114/204
    P(B) = 108/2014
    P(A and B) = 78/204
    P(A or B) = P(A) + P(B) - P(A and B)
              = 114/204 + 108/204 - 78/204
              = (114 + 108 - 78)/ 204
              = 144/204
              = 0.70588
```

(b)

```r
prob <- 78/ 114
prob
```

```
## [1] 0.6842105
```

(c)

```r
prob_male_brown_blue <- 19/ 54
prob_male_brown_blue
```

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

```r
prob_male_green_blue <- 11/36
prob_male_green_blue
```

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

(d)

```
No, the eye colors of male respondents and their partners are not independent.
Because the probability of same eye color of male and female has greater
probability than any other eye color.
```

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.
(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.
(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.
(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.

** Answer **

(a)

```
Prob_hardcover = 28/95
Prob_paperback_fiction = 59/94
(Note: Here we have already drawn one book so total number of book is 94)
```

```r
prob2 <- (28/95 * 59/94)
prob2
```

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

```
(b) prob_fiction = 72/95
    prob_hardcover = 28/94
```

```r
prob3 <- (72/95 * 28/94)
prob3
```

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

```
(c) prob_fiction = 72/95
    prob_hardcover = 28/95
```

```r
prob4 <- (72/95 * 28/95)
prob4
```

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

```
(d) In part (b) and (c) there is only change in 1 book. Because of that 1 book small probability
change occure in the scinario.
```

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.
(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.

** Answer **

(a)

```
bags <- c(0,1,2)
fees <- c(0, 25, 60)
probability <- c(0.54, 0.34, 0.12)
revenue <- fees * probability
revenue
```

```
## [1] 0.0 8.5 7.2
```

```
# below average is for one passenger revenue
avg_revenue <- 0.0+8.5+7.2
avg_revenue
```

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

```
variance <- (0 - 15.7) ^ 2 * 0.54 + (25 - 15.7) ^ 2 * 0.34 + (60 - 15.7) ^ 2 * 0.12
variance
```

```
## [1] 398.01
```

```
sd_revenue <- round(sqrt(variance),2)
sd_revenue
```

```
## [1] 19.95
```

```
(b) Number of passenger = 120
If revenue for 1 passenger is 15.7 then assuming for 120 passenger revenue will be revenue * 120.
```

```
avg_revenue120 <- 120 * avg_revenue
avg_revenue120
```

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

```
variance120 <- 120 * variance
sd_revenue120 <- round(sqrt(variance120),2)
sd_revenue120
```

```
## [1] 218.54
```

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.

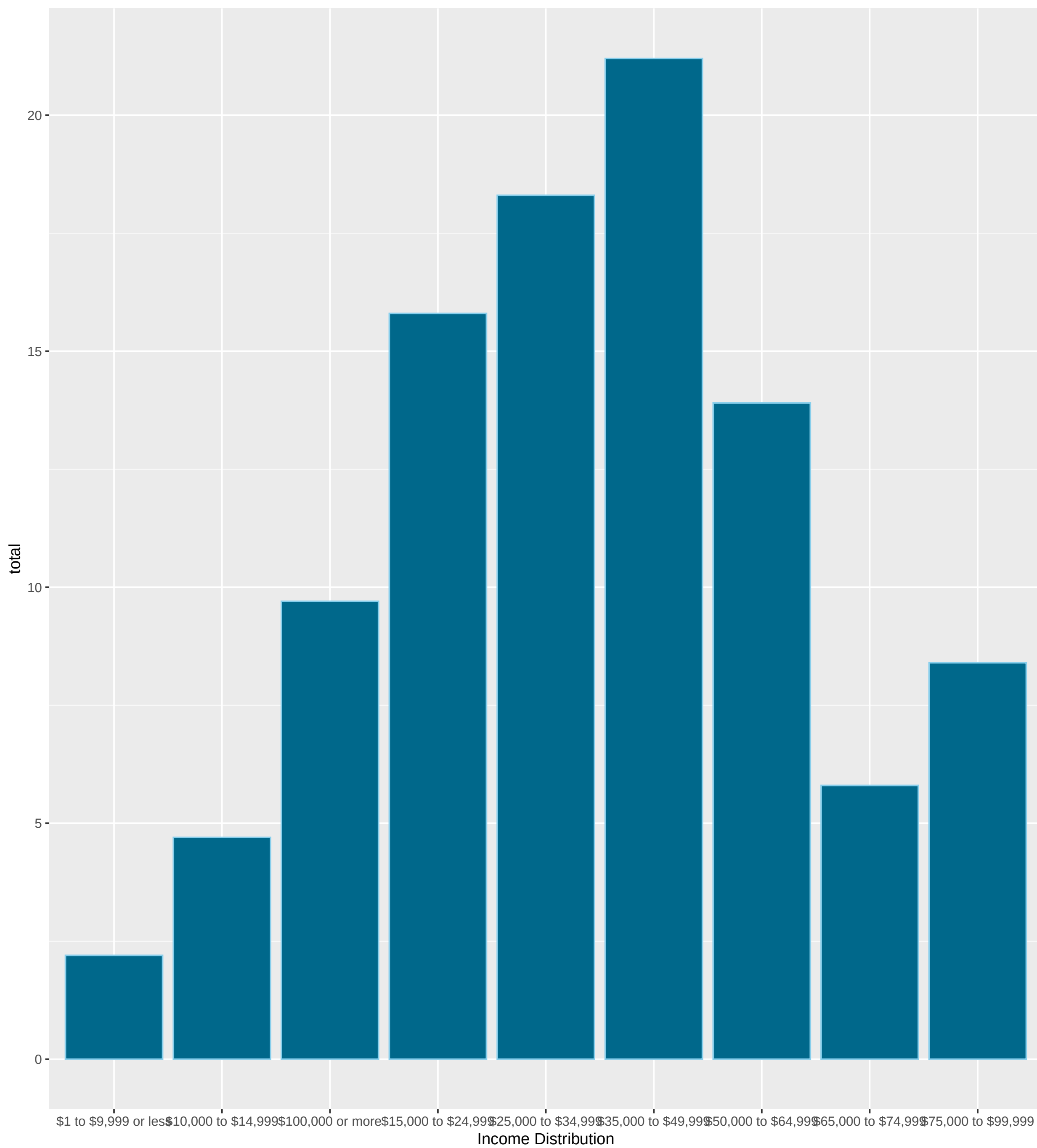
Income	*Total*
\$1 to \$9,999 or loss	2.2%
\$10,000 to \$14,999	4.7%
\$15,000 to \$24,999	15.8%
\$25,000 to \$34,999	18.3%
\$35,000 to \$49,999	21.2%
\$50,000 to \$64,999	13.9%
\$65,000 to \$74,999	5.8%
\$75,000 to \$99,999	8.4%
\$100,000 or more	9.7%

(a) Describe the distribution of total personal income.
(b) What is the probability that a randomly chosen US resident makes less than \$50,000 per year?
(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.
(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 **

(a) The distribution seems like unimodal and right skewed.

```
library(ggplot2)
data <- data.frame(income = c("$1 to $9,999 or less", "$10,000 to $14,999", "$15,000 to $24,999","$25,0
            "$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))
ggplot(data, aes(x=income, y=total)) + geom_bar(stat = "identity", color="sky blue", fill = "deepskyblu
```



(b) Probability that a randomly chosen US resident makes less than $50,000 per year
= 2.2 + 4.7 + 15.8 + 18.3 + 21.2 = 62.2

(c) Probability that a randomly chosen US resident makes less than $50,000 per year and is female
= US resident makes less than $50,000 per year * female %
= 62.2 * .41
= 25.502
Assuming that income is independent based on male and female.

(d) The same data source indicates that 71.8% of females make less than $50,000 per year
= 62.2 * 0.718

```
= 44.6596
My assumetion for (c) is wrong, after increse female pecent the income varies.
So, income is dependent on male and female.
```