Chapter 3 - Probability

Chunjie Nan

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: Not possible.

(b) getting a sum of 5?

Answer: (1,4) (4,1) (2,3) (3,2) so, there are 4 possible ways to get a sum of 5, and the total possible outcome of 2 pair dices are 6*6=36. Therefore, the probability of getting sum of 5 with 2 dices are $4/36 = 1/9$.

(c) getting a sum of 12?

Answer: The only way to get 12 is when both dices comes 6. so, 1/36 is the probability to get a sum of 12.

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: they are not disjoint events because 4.2% fall into both categories.

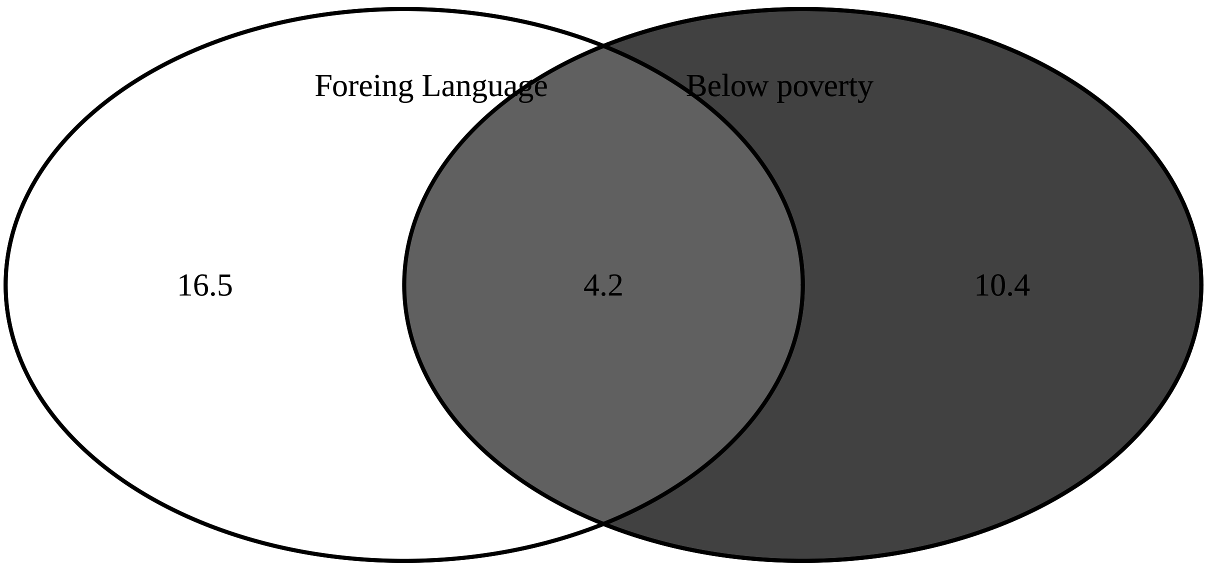
(b) Draw a Venn diagram summarizing the variables and their associated probabilities. Answer:

```
library(VennDiagram)
```

```
## Loading required package: grid

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

```
library(grid)
venn <- draw.pairwise.venn(14.6, 20.7, 4.2, c("Foreing Language       ", "        Below poverty"),  scal
grid.draw(venn)
```



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

Answer: According to the graph, it shows 10.4% of Americans live below the poverty who only speak English at home.

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

Answer: 14.6%+20.7%-4.2%= 31.1% of Americans live below the poverty line or speak a foreign language.

(e) What percent of Americans live above the poverty line and only speak English at home? 1-31.1%=68.9% of Americans live above the poverty line and speak only English.

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

They are not independent because some of the Americans fall into both below the poverty line and speak foreign language at home.

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?

Answer: Answer: 114 is total male has blue eyes, 19 is female has blue eyes whose husband has brown eyes 11 is female has blue eyes whose husband has green eyes 204 is total

```
(114+19+11)/204
```

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

Therefore, 70.59% of probability that randomly choose male or his wife have blue eyes.

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

Answer: 78/114 = 68.42% About 68.42% of male has blueeyes have a wife with blue eyes.

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

Answer: 19/54 = 35.19% About 35.19% of probability that a randomly chosen male with brown eyes has his wife have blue eyes. 11/35 = 30.56% About 30.56% of probability that a randomly chosen male with green eyes has his wife have blue eyes.

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

They are not independent since some wife have blue eyes not only married a husband who has blue eyes, but also a husband who has brown or green eyes as well.

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.

Answer: (28/95)*(59/94)=18.50% It has 18.50% of probability shows first drwaing a hardcover and then draw the paper fiction.

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

Answer: (72/95)*(28/94)= 22.58% About 22.58% of probability.

(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: (72/95)*(28/95)=22.34% It has 22.34% of probability.

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

Answer: I believe that putting back a book before draw the second book doesn't change the outcome that much becuase the overall samples on the book list are large enough to ignore the change of book quantity.

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.

Answer:

```
bag_0<-0
bag_1<-25
bag_2<-35
fee<-c(bag_0,bag_1,bag_2)
percentile<-c(0.54,0.34,0.12)
revenue<- sum(fee*percentile)
revenue
```

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

```
variance<- (25^2 *0.34 + 35^2 *0.12)-12.7^2
std<-sqrt(variance)
std
```

```
## [1] 14.07871
```

the revenue per passenger is \$12.7, and the standard deviation is 14.078.

(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: 12.7dollar * 120 = 1524 dollars

```
new_std<- sqrt(variance*120)
new_std
```

```
## [1] 154.2245
```

so, expected revenue for 120 passengers is 1524 dollars, and the standard deviation is 154.2245 dollars.

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.

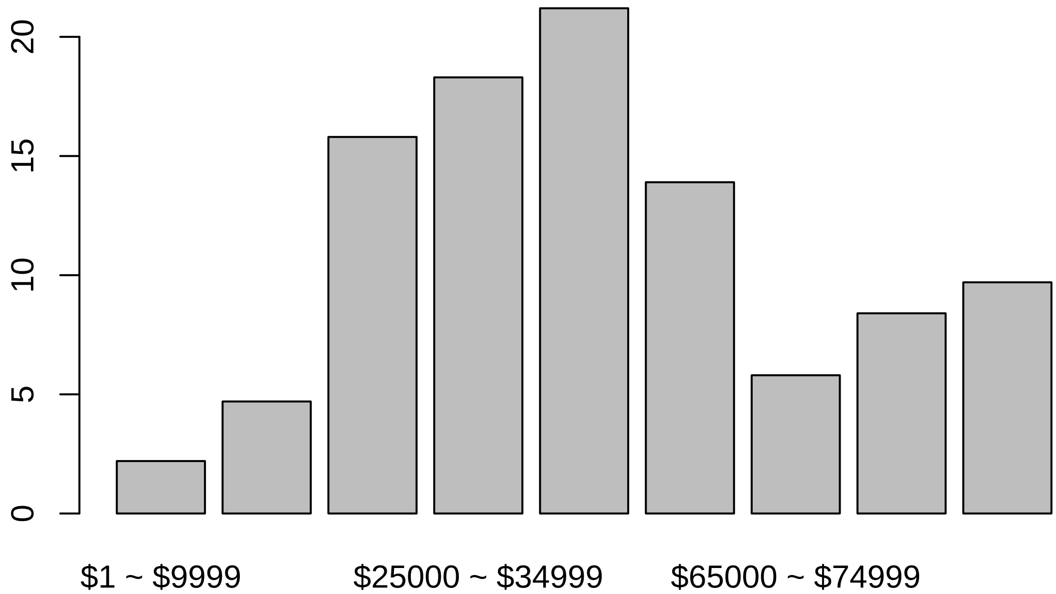
Answer:

```
personal_income<- c("$1 ~ $9999", "$10000 ~ $14999", "$15000 ~ $24999", "$25000 ~ $34999", "$35000 ~ $4
total <- c(2.2,4.7,15.8,18.3,21.2,13.9,5.8,8.4,9.7)

dist <- data.frame(personal_income, total)
dist
```

```
##    personal_income total
## 1       $1 ~ $9999   2.2
## 2  $10000 ~ $14999   4.7
## 3  $15000 ~ $24999  15.8
## 4  $25000 ~ $34999  18.3
## 5  $35000 ~ $49999  21.2
## 6  $50000 ~ $64999  13.9
## 7  $65000 ~ $74999   5.8
## 8  $75000 ~ $99999   8.4
## 9 $100000 or above   9.7
```

```
barplot(dist$total, names.arg = personal_income)
```



It shows a normal distribution according to the graph.

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

Answer: 2.2 + 4.7 + 15.8 + 18.3 + 21.2 = 62.2% About 62.2% 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?

Answer: 62.2% * 0.41=25.5% A 25.5% of US female makes less than $50,000 per year.

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: I conclude that our assumption is wrong if 71% of females makes less than $50,000. I believe it might be caused by the events are not independent in our assumption.