Chapter 3 - Probability

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?

***Answers: a) 0

(b) getting a sum of 5?

Roll	1	2	3	4
die 1	1	2	3	4
die 2	4	3	2	1

There are 4 outcomes which can result in a sum of 5, and 36 total outcomes possible (6 X 6), therefore the probability is $\frac{4}{36} = \frac{1}{9} \approx 0.1111111$

(c) getting a sum of 12? here is only one outcome from 2 dice which sum to 12: a 6 and 6 (boxcars). As such, the probablity is:

$\frac{1}{36} \approx 0.0277778$

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 : 4.2% of people fall under both category so it is not disjoint.******

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

```r
library(VennDiagram)
```

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

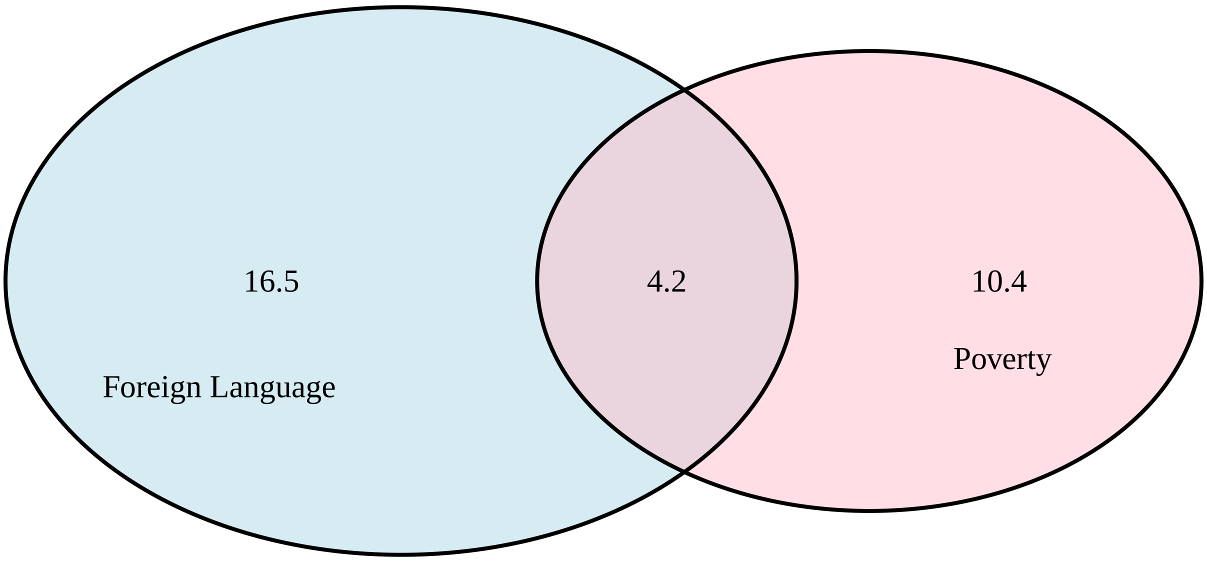
```
## Loading required package: futile.logger
```

```r
pov <- 14.6
forLang <- 20.7
both <- 4.2
povOnly <- pov - both
forLangOnly <- forLang - both

##ven_d <- draw.pairwise.venn(pov, forLang, both, c("Foreign", "Poverty"),  scale = FALSE, fill = c("gr

venn.plot <- draw.pairwise.venn(pov,
                   forLang,
                   cross.area=both,
                   c("Poverty", "Foreign Language"),
                   fill=c("pink", "lightblue"),
                   cat.dist=-0.08,
                   ind=FALSE)

grid.draw(venn.plot)
```



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

Based on the results of the Venn diagram, 10.4% of American 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? Using the General Addition Rule from the text, where P(A or B) = P(A) + P(B) - P(A and B), then: ***this would be P(A) + P(B) - p(A and B)***

$20.7 + 14.6 - 4.2 = 31.1\%$

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

79.3% English-speaking Americans - 10.4 English speaking Americans in Povery = 68.9% Americans living above 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?

Using the test of the Multiplication Rule for independent events (p91), do poverty and language satisfy the rule:

P(A and B) = P(A) x P(B)

$0.146 \times 0.207 = 0.030222$

This is not equal to 0.042, therefore it fails the Multiplication Rule for independent events test - the events are not independent and therefore knowing information about one of the events does provide information about the outcome of the other event. *

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 :

P(Male Blue or Partner Blue) $= \frac{114}{204} + \frac{108}{204} - \frac{78}{204} = 0.7058824$

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

P(Partner Blue given Male Blue) $= \frac{78}{114} = 0.6842105$ **************************

(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 :

P(Partner Blue given Male Brown) $= \frac{19}{54} = 0.3518519$

P(Partner Blue given Male Green) $= \frac{11}{36} = 0.3055556$ **************************

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

Answer The events are not independent because p(A INTERSECTION B) IS NOT EQUAL TO P(A)P(B) also Looking at proportions by male eye color, there is definitely an affinity to selecting a partner with the same 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.

Answer: P(Hardcover) $= \frac{28}{95} = 0.2947368$

Then the joint probability of paperback fiction (w/o replacement):

P(Paperback Fiction) $= \frac{59}{94} = 0.6276596$

P(Hardcover and Paperback Fiction) $= 0.2947368 \times 0.6276596 = 0.1849944$

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

Answer:

First we identify the marginal probability for Fiction:

P(Fiction) $= \frac{72}{95} = 0.7578947$

Then the marginal probability of hardcover fiction (w/o replacement):

P(Hardcover) $= \frac{28}{94} = 0.2978723$

P(Fiction and Hardcover) $= 0.7578947 \times 0.2978723 = 0.2257559$

(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.

Answer: We know the marginal probability for Fiction:

P(Fiction) $= \frac{72}{95} = 0.7578947$

But now the marginal probability of hardcover fiction is based on replacement:

P(Hardcover) $= \frac{28}{95} = 0.2947368$

P(Fiction and Hardcover) $= 0.7578947 \times 0.2947368 = 0.2233795$

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

Answer: from a large set of 95 books just 1 book is being replaced which does not cause much difference.

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. ** ##for 12% passengers cost to carry 2 bags would 25+35=60

```
no_of_bags<-c(0,1,2)
fees<-c(0,25,60)
passengers<-c(.54,.34,.12)

baggagedata<-data.frame(no_of_bags,fees,passengers)
baggagedata
```

```
##   no_of_bags fees passengers
## 1          0    0       0.54
## 2          1   25       0.34
## 3          2   60       0.12
```

```
average=sum(baggagedata$passengers*baggagedata$fees)/sum(baggagedata$passengers)

sprintf("Average revenue per passengers is %s", average)
```

```
## [1] "Average revenue per passengers is 15.7"
```

```
standarddeviation <- sqrt(0.54*(0-average)^2 + 0.34*(25-average)^2 + 0.12*(60-average)^2)
sprintf("Standard deviation is %s", standarddeviation)
```

```
## [1] "Standard deviation is 19.9501879690393"
```

(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: ##Based on the input given for the case of 120 passengers proportion would be 65 have 0 checked in bags, 41 have 1 checked in bag and 14 have 2 checked in bags.

```
revenuefrom_120 <- (65*0 + 41*25 + 14 * 60)
sprintf("Revenue from 120  passengers will be  %s", revenuefrom_120)
```

```
## [1] "Revenue from 120  passengers will be  1865"
```

```
avgof_120 <- revenuefrom_120/120
standard_deviation_120 <- sqrt(0.54*(0-avgof_120)^2 + 0.34*(25-avgof_120)^2 + 0.12*(60-avgof_120)^2)

sprintf("Standard deviation for   120  passengers will be  %s", standard_deviation_120)
```

```
## [1] "Standard deviation for   120  passengers will be  19.9508162601044"
```

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%

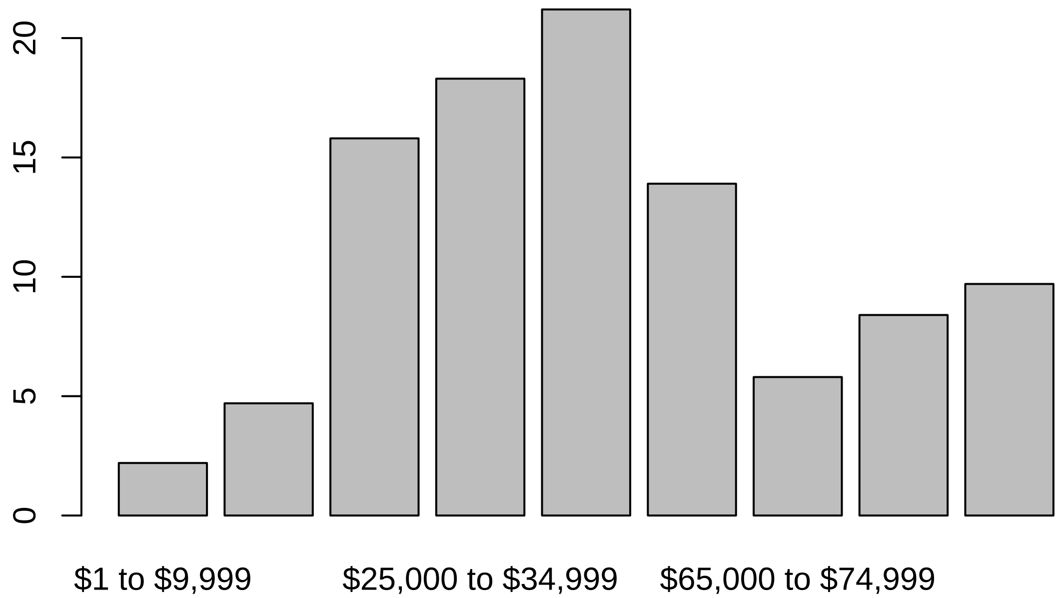
(a) Describe the distribution of total personal income.

```
distribution <- data.frame(
income <- c("$1 to $9,999","$10,000 to $14,999","$15,000 to $24,999","$25,000 to $34,999","$35,000 to $4
total <- c(2.2,4.7,15.8,18.3,21.2,13.9,5.8,8.4,9.7))

distribution
```

```
##   income....c...1.to..9.999.....10.000.to..14.999.....15.000.to..24.999...
## 1                                                             $1 to $9,999
## 2                                                       $10,000 to $14,999
## 3                                                       $15,000 to $24,999
## 4                                                       $25,000 to $34,999
## 5                                                       $35,000 to $49,999
## 6                                                       $50,000 to $64,999
## 7                                                       $65,000 to $74,999
## 8                                                       $75,000 to $99,999
## 9                                                         $100,000 or more
##   total....c.2.2..4.7..15.8..18.3..21.2..13.9..5.8..8.4..9.7.
## 1                                                         2.2
## 2                                                         4.7
## 3                                                        15.8
## 4                                                        18.3
## 5                                                        21.2
## 6                                                        13.9
## 7                                                         5.8
## 8                                                         8.4
## 9                                                         9.7
```

```
barplot(distribution$total, names.arg=income)
```



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

```
probMakingLessthan50K <- sum(distribution[1:5,]$total)
probMakingLessthan50K
```

```
## [1] 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 income and gender are independent events

```
P_female <- 0.41
P_50K <- probMakingLessthan50K
# Compute
P_female_50K <- P_female * P_50K
# Show the result
P_female_50K
```

```
## [1] 25.502
```

(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.

```r
# What percent of population is female and makes less than $50K?
actualFemale50K <- 0.718 * P_female
actualFemale50K
```

```
## [1] 0.29438
```

While 0.29438 is close to 25.502, they are not equal. Therefore I conclude that making less than $50K and being female are not independent events.