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?
**This is not possible since all summations will be 2 or greater

(b) getting a sum of 5?
Possibilities:
Die 1: roll a 1 Die 2: roll a 4
Die 1: roll a 2 Die 2: roll a 3
Die 1: roll a 3 Die 2: roll a 2
Die 1: roll a 4 Die 2: roll a 1

OR can be viewed as a 1/6 chance per die to select 1 of 4 options so

```
#x<-(4*(1/6*1/6))
percent(4*(1/6*1/6))
```

```
## [1] "11%"
```

or more accurately 11.11%

(c) getting a sum of 12?

Possibilities: Die 1: roll a 1 Die 2: roll a 4
or 1 * (1/6 * 1/6) = 1/36 = 27.78%

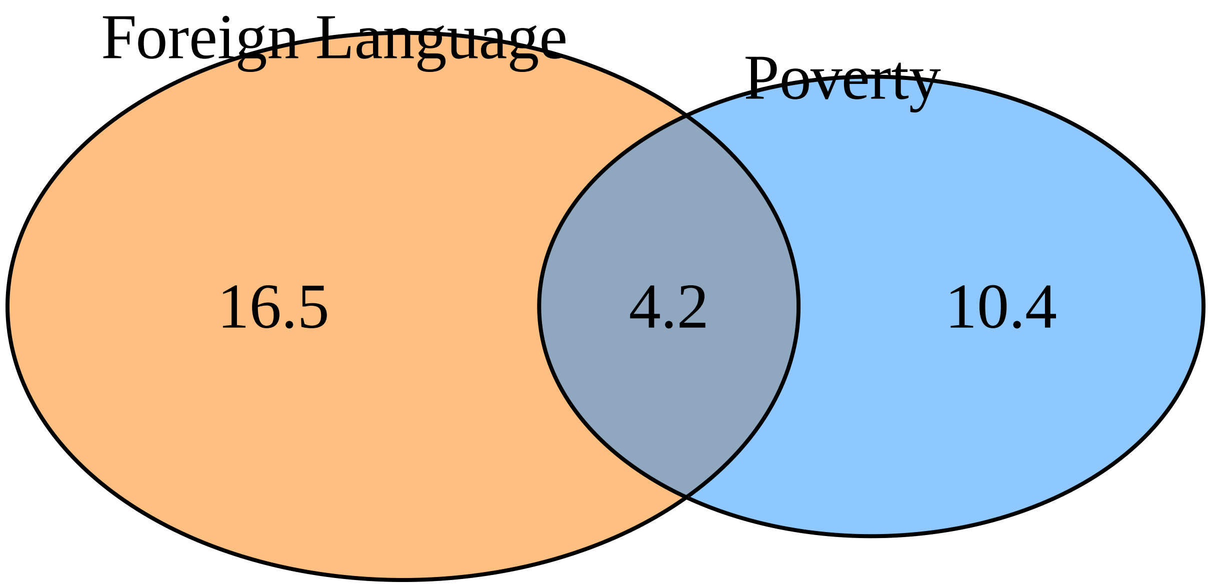
```
1/6*1/6
```

```
## [1] 0.02777778
```

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?
In order to qualify as disjoint each event must be mutually exclusive.
P(BP) = Probability of being below the poverty line = 14.6%
P(FL) = Probability of that the American speaks a Foreign Language = 20.7%
P(BP and FL) = 4.2% therefore it cannot be disjoint

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




```
## (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?
P(BP)-P(BP and FL)= 14.6 - 4.2 = 10.4
or simply note the Poverty region shaded in blue.

(d) What percent of Americans live below the poverty line or speak a foreign language at home?
P(BP | FL) = (P(BP) + P(FL)) - P(BP and FL)=

```r
(14.6 + 20.7) - 4.2
```

```
## [1] 31.1
```

31.1%

(e) What percent of Americans live above the poverty line and only speak English at home?
Above calculations account for population that speaks a foreigh language and for the population below the poverty line. Therefore the percent above the poverty line that speaks English is 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?
**P(A and B) = P(A) * P(B)

```r
.146*.207
```

```
## [1] 0.030222
```

not equal to 4.2% or .042 therefore not 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? **(P(Male_Blue)+P(Female_Blue))-P(Both_blue)**

```
(114/204+108/204)-78/204
```

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

(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes? **P(Male_Blue AND Female_Blue) / P(Male_blue)
(78/204) / (114/204) = 78/114

```
(78/204)/(114/204)
```

```
## [1] 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?

P(male_brow and women_blue) / P(Male with brown eyes) = (19/204) / (54/204) = 19/54

```
(19/204) / (54/204)
```

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

P(male_green and women_blue) / P(male_green) = (11/204) / (36/204) = 11/36

```
(11/204) / (36/204)
```

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

(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning. **No, if they were independent th probability of Females with blue eyes would equal the probability of female with blue eyes and males with blue eyes (78/14 = 108/204) which it clearly does not.**

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.
P(Hardcover) = 28 / 95
P(Paperback and fiction and no replacement)59 / 94
*= 18.5%**

```
(28/95)*(59/94)
```

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

(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement. **P(fiction_hardcover_first)= 13/95 then P(hardcover_no_replace) = 27/94**
or P(fiction_softcover_first)=59/95 then P(hardcover_no_replace) = 28/94
Then add the two probabilities.

```
(13/95*27/94)+(59/95*28/94)
```

```
## [1] 0.2243001
```

(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. **P(fiction)=72/95 P(hardcover)=28/95

```
72/95*28/95
```

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

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

The sample size is relatively large and removing on book from the equation (denominator or w/o replace) is not very significant.

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.

```
# set possible number of luggage
Pieces_of_luggage<-c(0,1,2)
#set fee vector NOTE:2 baggages 25+35=60
Fees<-c(0,25,60)
#set percentage of luggage buy customer
Percentage_of_Pieces<-c(0.54, 0.34, 0.12)
#create dataframe
Baggage_Fees_df<-data.frame(Pieces_of_luggage,Fees,Percentage_of_Pieces)
#Expected value E(x) is the summation of the fees by there percentages
Avg_Rev <- sum(Fees * Percentage_of_Pieces)
#show data
#Baggage_Fees_df
Avg_Rev
```

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

```
Stand_Dev<-sqrt((0 - Avg_Rev)^ 2 * 0.54 + (25 - Avg_Rev)^ 2 * 0.34 + (60 - Avg_Rev)^ 2 * 0.12)
Stand_Dev
```

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

Average per passenger is 15.7 and Standard Deviation is 19.95

(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.

The airline can expect $1884 from a flight of 120 passengers and a standard deviation of 281.54

```
120*15.7
```

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

```
Stand_Dev_120<-sqrt((0 - Avg_Rev)^ 2 * 0.54 * 120 + (25 - Avg_Rev)^ 2 * 0.34 * 120 + (60 - Avg_Rev)^ 2 *
Stand_Dev_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.

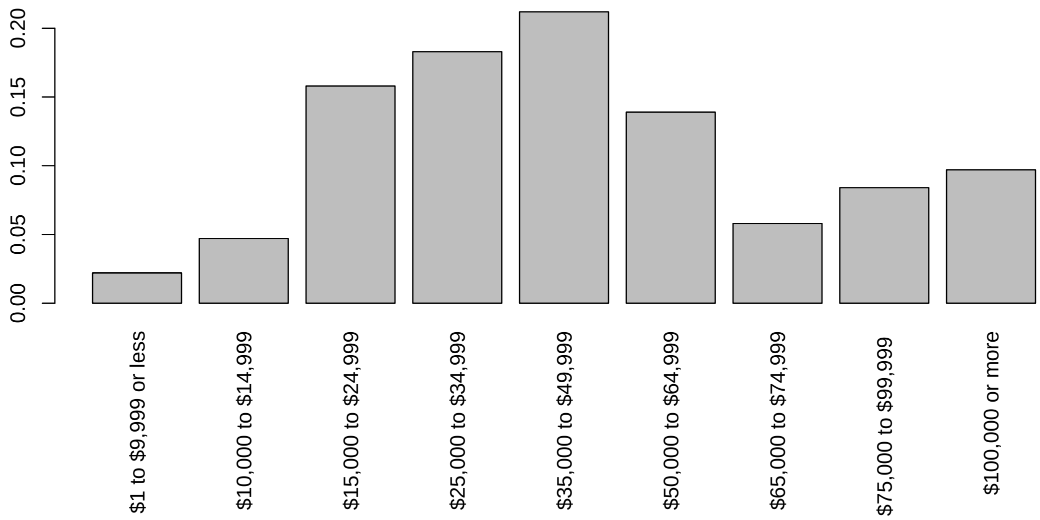
Income	*Total*
$1 to $9,999 or loss	2.2%
$10,000 to $14,999	4.7%
$15,000 to $24,999	15.8%
$25,000 to $34,999	18.3%
$35,000 to $49,999	21.2%
$50,000 to $64,999	13.9%
$65,000 to $74,999	5.8%
$75,000 to $99,999	8.4%
$100,000 or more	9.7%

(a) Describe the distribution of total personal income.
Creating a data frame for the variables Income and Total is needed
The distribution is bimodal with 2 peaks

```
# store vector income
Income <- c("$1 to $9,999 or less", "$10,000 to $14,999", "$15,000 to $24,999", "$25,000 to $34,999", "$
# store vector total, decimals used for easy calculations and because percent() function removes decima
Total <- c(0.022, 0.047, 0.158, 0.183, 0.212, 0.139, 0.058, 0.084, 0.097)
# create dataframe Income_Gender
Income_Gender_df <- data.frame(Income, Total)
# show data
Income_Gender_df
```

```
##                 Income Total
## 1 $1 to $9,999 or less 0.022
## 2   $10,000 to $14,999 0.047
## 3   $15,000 to $24,999 0.158
## 4   $25,000 to $34,999 0.183
## 5   $35,000 to $49,999 0.212
## 6   $50,000 to $64,999 0.139
## 7   $65,000 to $74,999 0.058
## 8  $75,000 to $99,999  0.084
## 9     $100,000 or more 0.097
```

```
par(mar=c(9,4,4,5)+.1)
barplot(names.arg = Income, Total, las = 3)
```



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

```
# sum values below '$50,000 to $64,999'
0.022+0.047+0.158+0.183+0.212
```

```
## [1] 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.

```
# above results times the probability of being a female in this sample
0.622*.41
```

```
## [1] 0.25502
```

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.

25.5% != 71.8% Therefore not independent