Data 606 Chapter 3 - Probability

Homework Completed by Chad Bailey

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?

```
##  1.a The probability of rolling a sum of 1 is 0%. The minimum observable sum is 2 {}
##  1.b The probability of rolling a sum of 5 is 1/9 (approimately 11.11%).
##      {(1, 4) (2,3), (3, 2), (4, 1)}
##  1.c The probability of rolling a sum of 12 is 1/36 (approximately 2.78%) {(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?

```
##  2.a Living below the poverty line and speaking a foreign language at home are
##      NOT disjoint. It is fully possible to live below poverty and also speak a
##      foreign language.
##  2.b
##  2.c The percent of Americans living below the poverity line and only speak English
##      at home is P(Below_Poverty) - P(Both) = 14.6% - 4.2% = 10.4%
##  2.d The percent of Americans living below the poverty line or speak a foreign
##      language at home is P(Below_Poverty) + P(Foreign_Language) - P(Both)
##      = 14.6% + 20.7% - 4.2% = 31.1%
##  2.e The percent of Americans living above the poverty line and only speak English
##      at home is 100 - (P(Below_Poverty) + P(Foreign_Language) - P(Both))
##      = 100% - 31.1% = 68.9%
##  2.f Using the Mulitiplication Rule for Independent Processees, we can show that
##      unfortunately, no, the event that someone lives below the poverty line is
##      independent of the event that the person speaks a foreign language at home
##      are NOT independent.
##
##      If the events were independent then
##      P(Poverty AND Language) = P(Poverty) x P(Language)

##      It was given that
##      P(Poverty AND Language) = 4.2%
##      P(Poverty) = 14.6%
##      P(Language) = 20.7%

##      P(Poverty) x P(Language) = 14.6% x 20.7% = 3.02% which is NOT equal to 4.2%.
##      Thus the two events are not fully 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.

```
##  3.a The probability  a randomly chosen male respondent or his partner has blue eyes
##      can be found by P(male_blue) + P(partner_blue) - P(both)
##      = 114/204 + 108/204 - 78/204 = 144/204 = 70.59%

##  3.b The probability a randomly chosen male respondent with blue eyes has a partner
##      with blue eyes is 78/114 = 68.42%.

##  3.c The probability a randomly chosen male respondent with brown eyes has a partner
##      with blue eyes is 19/54 = 35.19%

##      The probability a randomly chosen male respondent with green eyes has a partner
##      with blue eyes is 11/36 = 30.56%

##  3.d No, it does not appear that eye colors of male respondents and their partners
##      are independent. The distribution of observed pairs does not follow the
##      Mulitiplication Rule for Independent Processees.

##      The probablity of male having blue is 114/204 = 56%
##      The probablity of partner haing blue is 108/204 = 53%
##      The observed probablity of both male and partner blue is 78/204 = 38%.
##      P(male_blue) x P(partner_blue) = 56% x 53% = 30%
##      Since P(male_blue) x P(partner_blue) is NOT equal to P(male_blue AND partner_blue)
##      then these events are not independent.
```

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.

```
## 4.a P(hardcover) = 28/95 = 29.5%
##     P(paperback fiction without replacement) = 59/94 = 62.8%
##     P(hardcover then paperback fiction without replacement)
##     = 28/95 x 59/94 = 1652/8930 = 18.5%


## 4.c P(fiction) = 72/95 =83.2%
##     P(hardcover) = 28/95 = 29.5%
##     P(fiction then hardcover) = 72/95 x 28/95 = 2016/9025 = 22.3%


## 4.d The results are very similar because the scenarios differ by only one book.
##     The results would differ more if the number of selections was greater.
```

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.

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.

```
##  6.a The distribution is right skewed
##  6.b The probability that a randomly chosen US resident makes less than $50,000
##      per year is 62.2%
##  6.c The probability that a randomly chosen US resident makes less than $50,000
##      per year and is female is 62.2% x 41% = 25.5%
##      Assuming income is independent of gender
##  6.d Given that 71.8% of females make less than $50,000, it is clear the assumption
##      that income is independent of gener is invalid.
```