Data606 Week 2

Sedo

September 2, 2017

2.6 Dice Rolls

a) getting a sum of 1? 0
b) getting a sum of 5? 4/36
c) getting a sum of 12? 1/36

2.8 Poverty and Language

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?

No, people can be both below the poverty line and speak foreign language at home.

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

```
library(VennDiagram)
```

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

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

```
venn <- draw.pairwise.venn(area1=.146, area2 = .207, cross.area = .042, c("Poverty", "Foreign"))
grid.draw(venn)
```
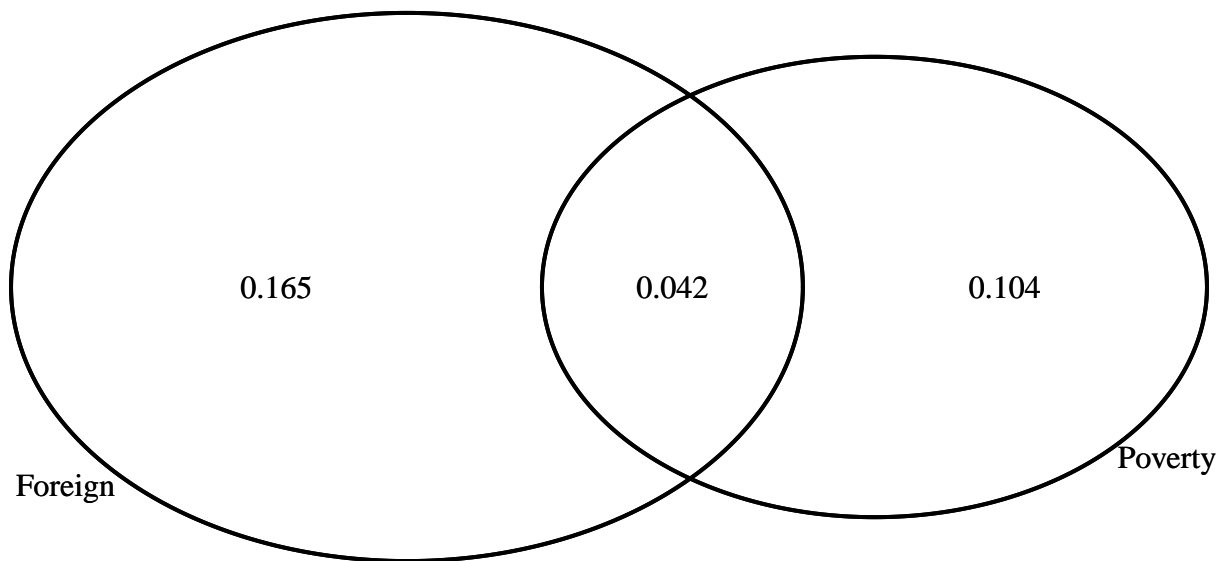
```
grid.newpage()
```

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

10.4%

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

4.2%

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

68.9%

```
print(1-0.165-0.042-.104)
```

```
## [1] 0.689
```

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

Probabilty of someone poverty = 14.6% Probabilty of someone poverty given foreign language = 20.3%

The difference is decently large at 5.7%. There is a high chance that poverty dependent on person speaking foreign language at home.

```
print(0.042/0.207)
```

```
## [1] 0.2028986
```

2.20 Assortative mating.

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.

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

70.6%

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

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

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

68.4%

```
print(78/114)
```

```
## [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?

35% for brown eyes partner with blue 31% for green eyes partner with blue

```
print(19/54)
```

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

```
print(11/36)
```

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

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

No. Blue eye males have almost twice as likely chance of partnering with blue eye females than Brown or Green.

2.30 Books on a bookshelf

The table below shows the distribution of books on a bookcase based on whether they are nonfiction or fiction and hardcover or paperback.

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.

18.5%

```
print((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.

22.4%

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

```
## [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.

22.4%

```
print(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 answers are similar because removing just one book out of 95 books does not much difference to the set whether there's replacement or not.

2.38 Baggage fees

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.

```
average = (0.54*0 + 0.34*25 + 0.12*60)
print(average)
```

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

```
var = (0-average)^2*0.54+(0.34*25-average)^2*0.34+(0.12*60-average)^2*0.12
sd=sqrt(var)
print(sd)
```

```
## [1] 12.62538
```

   (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

$1884 is expected amount. 1515 is the standard deviation.

Assumption is that a passenger's decision to take on any number of bags is independent of another passenger's decision to take any number of bags.

```
print(120*average)
```

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

```
print(sqrt(120^2*var))
```

```
## [1] 1515.046
```

2.44 Income and Gender

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.

(a) Describe the distribution of total personal income.

The income is skewed to the right. Median income is between 35K to 50K.

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

62.2%

```r
print(2.2+4.7+15.8+18.3+21.2)
```

```
## [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.

25.5%, assuming at any given income level, the disritbution across male and female remains the same.

```r
print(62.2*.41)
```

```
## [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.

Since my assumption was that the distribution of male and female remained similar across the income levels, the fact that 71.8% makes less than 50K would suggest that the distribution does not remain similar. If it was similar, we would be expecting 62.2% of female to make less than 50K.