Rolling Dice.

Poverty and Language

Associative Mating

Books on a bookshelf.

Baggage fees.

Income and gender.

DS606-HW3

Code

George Cruz

2020-09-13

Hide

```
library(tidyverse)
library(openintro)
```

Rolling Dice.

If you roll a pair of fair dice. What is the probability of getting a sum of:

a. 1= The probability is 0. As the lowest possible sum is 2.
b. 5= 4/36 as there are 4 possible combinations that will add up to 5.
c. 12= 1/36 as there is only 1 possible combination that will add up to 12.

Poverty and Language

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? *No, we can see 4.2% of Americans do fall into both categories*

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

Hide

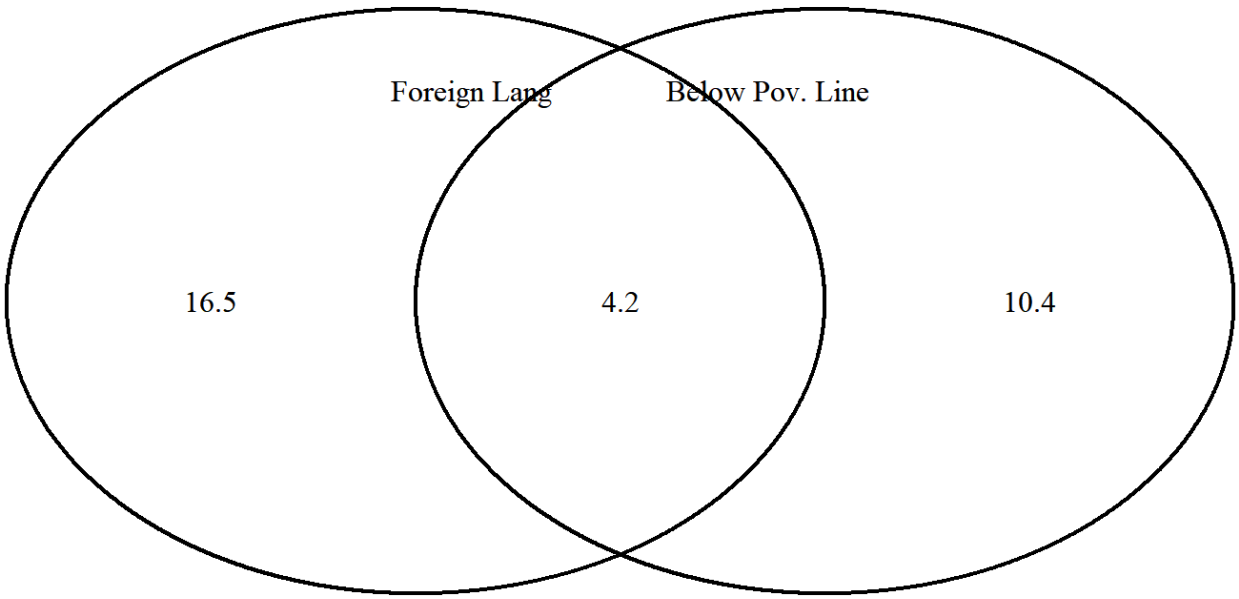
```
library(VennDiagram)
```

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

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

Hide

```
# A simple two-set diagram
venn.plot <- draw.pairwise.venn(14.6, 20.7, 4.2, c("Foreign Lang", "Below Pov. Line"),
         scaled=FALSE)
grid.draw(venn.plot)
```



Hide

```
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?

Hide

```
res <- 20.7 + 10.4
res
```

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

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

Hide

```
100 - 31.1
```

```
## [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? Probability of English Speaker to be bpl = 10.4/79.3 = 0.1311 Probability of Non-English = 4.2/20.7 = 0.2029

Hide

```
10.4/79.3
```

```
## [1] 0.1311475
```

Hide

```
4.2/20.7
```

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

Since there is a higher probability for a foreign language speaker to be below the poverty line, we can determine that these events are not independent.

Associative 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?

Hide

```
FemaleBlue <- 108/204
MaleBlue <- 114/204
bothBlue <- 78/204

ForMBlue <- FemaleBlue + MaleBlue - bothBlue
ForMBlue*100
```

```
## [1] 70.58824
```

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

Hide

```
(bothBlue/MaleBlue )*100
```

```
## [1] 68.42105
```

c. What is the probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes?

Hide

```
MaleBrown <- 54/204
femBlueMBrown <- 19/204
(femBlueMBrown/MaleBrown) * 100
```

```
## [1] 35.18519
```

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

Hide

```
MaleGreen <- 36/204
femBlueMGreen <- 11/204
(femBlueMGreen/MaleGreen) * 100
```

```
## [1] 30.55556
```

d. Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning. *Since there is a higher probability of matching eye colors, we can determine these events are dependent.*

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.

Hide

```
total <- 95
HcFiction <- 13
HcNon <- 15
PbFiction <- 59
PbNon <- 8
fiction <- 72
nonfic <- 23
HardCov <- 28
Paperb <- 67
```

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

Hide

```
(HardCov/total)*(PbFiction/(total-1)) * 100
```

```
## [1] 18.49944
```

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

Hide

```
(fiction/total)*(HardCov/(total-1))*100
```

```
## [1] 22.57559
```

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.

Hide

```
(fiction/total)*(HardCov/total)*100
```

```
## [1] 22.33795
```

d. The final answers to parts (b) and (c) are very similar. Explain why this is the case. *The probability is not significantly changed by removing one book because the total of books 95 is too large*

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.

Hide

```
baggageFees <- c(0, 25, 60)
passengers <- c(0.54, 0.34, 0.12)
probModel <- baggageFees * passengers

result <- data.frame("Avg"=mean(probModel), "SD"=sd(probModel))
result
```

```
##        Avg       SD
## 1 5.233333 4.578573
```

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.

Hide

```
data.frame( "Expected"=(120*sum(probModel)), "With_SD"=(120*result$SD))
```

```
##   Expected  With_SD
## 1     1884 549.4288
```

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.

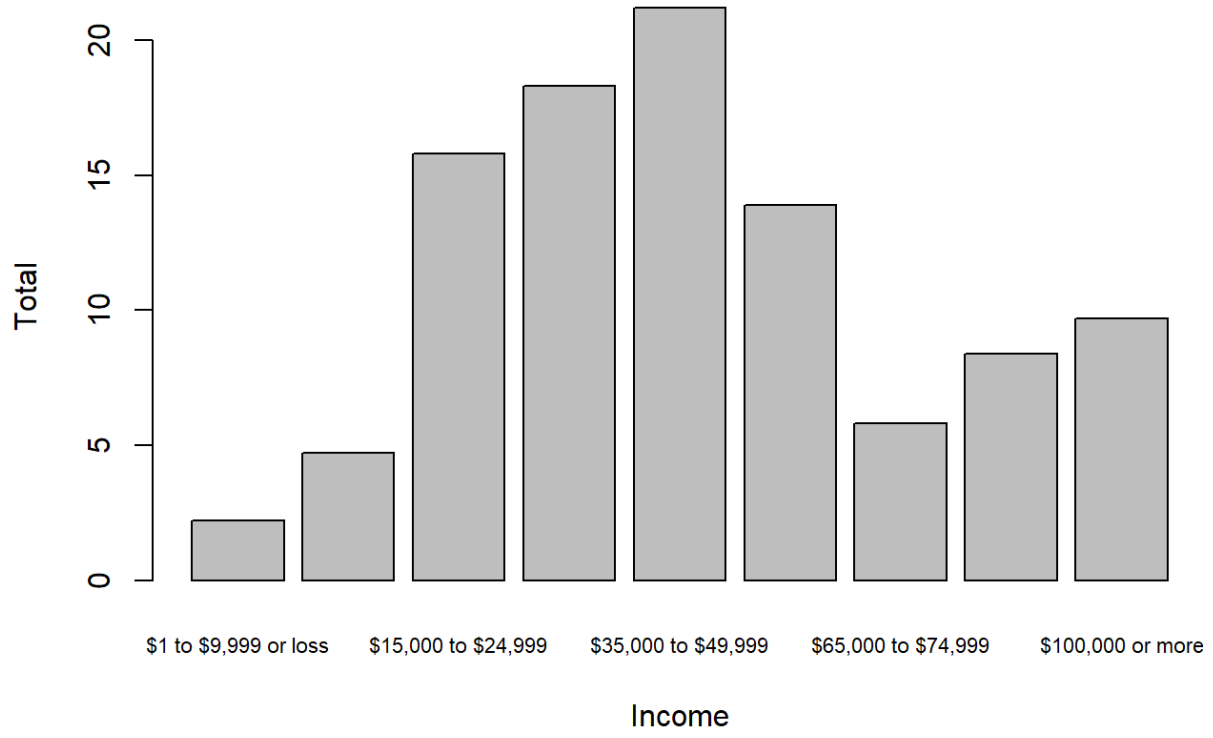
Hide

```
income <- c("$1 to $9,999 or loss","$10,000 to $14,999","$15,000 to $24,999","$25,000
          to $34,999","$35,000 to $49,999","$50,000 to $64,999","$65,000 to $74,999",
         "$75,000 to $99,999","$100,000 or more")
total <- c(2.2,4.7,15.8,18.3,21.2,13.9,5.8,8.4,9.7)
incomeGender <- data.frame("Income"=income, "Total"=total)
incomeGender
```

```
##                 Income Total
## 1 $1 to $9,999 or loss   2.2
## 2   $10,000 to $14,999   4.7
## 3   $15,000 to $24,999  15.8
## 4   $25,000 to $34,999  18.3
## 5   $35,000 to $49,999  21.2
## 6   $50,000 to $64,999  13.9
## 7   $65,000 to $74,999   5.8
## 8   $75,000 to $99,999   8.4
## 9     $100,000 or more   9.7
```

Hide

```
barplot(incomeGender$Total, names.arg = incomeGender$Income, xlab = "Income", ylab =
         "Total", cex.names=.7)
```



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

Hide

```
below50k <- sum(2.2,4.7,15.8,18.3,21.2)
(below50k/100)*100
```

```
## [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.

Hide

```
(below50k/100)*100*0.41
```

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

Assuming that salaries are evenly distributed among the population

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. *My Assumption was invalid and it appears salaries are not independent of gender*