Chapter 2 - Probability

CUNY MSDA - IS606 - Homework 2

Completed by: Duubar Villalobos Jimenez mydvtech@gmail.com

February 12, 2017

Homework

OpenIntro Statistics

```
Practice: 2.5, 2.7, 2.19, 2.29, 2.43
Graded: 2.6, 2.8, 2.20, 2.30, 2.38, 2.44
```

2.6 Dice rolls.

Based on a probability distribution table; we have as follows:

```
dice_2_sum <- c(2,3,4,5,6,7,8,9,10,11,12)
diece_prb <- c(1/36,2/36,3/36,4/36,5/36,6/36,5/36,4/36,3/36,2/36,1/36)
prob <- data.frame(dice_2_sum,diece_prb)
prob
```

```
##    dice_2_sum  diece_prb
## 1           2 0.02777778
## 2           3 0.05555556
## 3           4 0.08333333
## 4           5 0.11111111
## 5           6 0.13888889
## 6           7 0.16666667
## 7           8 0.13888889
## 8           9 0.11111111
## 9          10 0.08333333
## 10         11 0.05555556
## 11         12 0.02777778
```

If you roll a pair of fair dice, what is the probability of

(a) getting a sum of 1?

P(X=0) = 0

The probability of having a sum of 5 is 0%.

(b) getting a sum of 5?

P(X=5) = 4/36.

The probability of having a sum of 5 is 11.11%.

(c) getting a sum of 12?

P(X=12) = 1/36

The probability of having a sum of 5 is 2.78%.

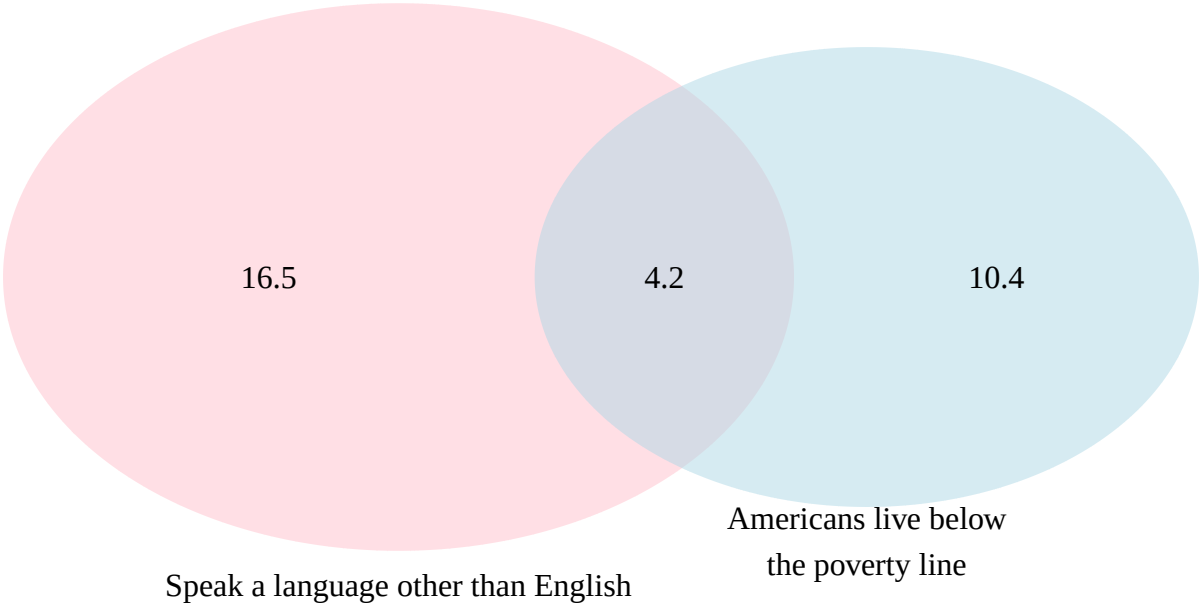
2.8 Poverty and language.

(a) Are living below the poverty line and speaking a foreign language at home disjoint?

No. They are not disjoint.

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

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




```
## (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(A live poverty | Only English at home) = P(A live poverty) - P(Speak other Language and A live poverty)

P(A live poverty | Only English at home) = 14.6 - 4.2

P(A live poverty | Only English at home) = 10.4

Answer: 10.4% 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?

P(A live poverty OR foreign at home) = P(A live poverty) + P(Speak other Language) - P(A live poverty and speak other language)

P(A live poverty | or foreign at home) = 14.6 + 20.7 - 4.2

P(A live poverty | or foreign at home) = 31.1

Answer: 31.1% 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?

P(Above poverty line) = 1 - (P(Speak other language) + P(A live poverty) - P(intersection))

```
100 - (20.7 + 14.6 - 4.2)
```

```
## [1] 68.9
```

Answer: 68.9% 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?

Not independent.

P(A live poverty and Speak other language) $\neq$ P(A live poverty and Speak) x P(Speak other language)

2.20 Assortative mating.

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

F = Female Blue, M = Male Blue

P (F) = 108 / 204 = 0.5294

P (M) = 114 / 204 = 0.5588

P (M $\cap$ F) = 78/204 = 0.38235

P (M $\cup$ F) = P (M) + P (F) - P (M $\cap$ F) = 108/204 + 114/204 - 78/ 204 = 144/204 = 0.7059

Answer: The probability will be 70.59%

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

P (F|M) = P (F $\cap$ M) / P (M) = (78 / 204) / (114 / 204) = 0.6842

Answer: the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes is 68.42%.

(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 (Female Blue Eyes | Male Brown Eyes) = $\frac{19}{54}$ = 0.3519 P (Female Blue Eyes | Male Green Eyes) = $\frac{11}{36}$ = 0.3056

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

The events are not independent because the probabilities are not the same.

If we analize as follows: P (Female Blue | Male Blue) 78 / 114 does not equal P (Female Blue) 108 / 204.

We can conclude that Probability of blue male and female eyes are not equal to female blue eyes, hense are not independent.

2.30 Books on a bookshelf

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

Lets define:

H: Hardcover

P: Paperback

Probability for Hardcover = $\frac{28}{95}$

Probability for Paperback (w/o replacement) = $\frac{59}{94}$

P = P(H) * P(P)

```
round(28/95 * 59/94,4)
```

```
## [1] 0.185
```

Answer: the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement is 18.50%

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

Probability for Fiction = $\frac{72}{95}$

Probability of hardcover fiction (w/o replacement) = $\frac{28}{94}$

```
round(72/95 * 28/94,4)
```

```
## [1] 0.2258
```

Answer: The probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement is 22.58%

(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.

Probability for Fiction = $\frac{72}{95}$

Probability of hardcover fiction with replacement:

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

```
round(72/95 * 28/95,4)
```

```
## [1] 0.2234
```

Answer: 22.34% with replacement.

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

The answers are very similar due the "high" number of books, one book does not make a big difference when we replace it or not.

2.38 Baggage fees.

(a) Build a probability model, compute the average revenue per passenger, and compute the corresponding standard deviation.

```
prob <- c(0.54, 0.34, 0.12)
bags <- c(0, 1, 2)
fees <- c(0, 25, 25 + 35)
my.data  <- data.frame(bags, fees, prob)

# Run a calculation for the revenue
my.data$Rev <-my.data$prob * my.data$fees
my.data
```

```
##   bags fees prob Rev
## 1    0    0 0.54 0.0
## 2    1   25 0.34 8.5
## 3    2   60 0.12 7.2
```

```
# Find the average revenue per passenger
avgRevPerPax <- sum(my.data$Rev)
avgRevPerPax
```

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

```
# Find the Variance
my.data$DiffMean <- my.data$Rev - avgRevPerPax
my.data$DiffMeanSqrd <- my.data$DiffMean ^ 2
my.data$DiffMeanSqrdTimesProb <- my.data$DiffMeanSqrd * my.data$prob
my.data
```

```
##   bags fees prob Rev DiffMean DiffMeanSqrd DiffMeanSqrdTimesProb
## 1    0    0 0.54 0.0    -15.7       246.49              133.1046
## 2    1   25 0.34 8.5     -7.2        51.84               17.6256
## 3    2   60 0.12 7.2     -8.5        72.25                8.6700
```

```
# Find the standard deviation
varRevPerPax <- sum(my.data$DiffMeanSqrdTimesProb)
sdRevPerPax <- sqrt(varRevPerPax)
sdRevPerPax
```

```
## [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.