Homework 3 - Probability

Peter Gatica

02/21/2021

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(DATA606)
```

```
## Loading required package: shiny

## Loading required package: openintro

## Loading required package: airports

## Loading required package: cherryblossom

## Loading required package: usdata

## Loading required package: OIdata

## Loading required package: RCurl

##
## Attaching package: 'RCurl'

## The following object is masked from 'package:tidyr':
##
##     complete

## Loading required package: maps
```

```
##
## Attaching package: 'maps'

## The following object is masked from 'package:purrr':
##
##     map

## Loading required package: markdown

##
## Welcome to CUNY DATA606 Statistics and Probability for Data Analytics
## This package is designed to support this course. The text book used
## is OpenIntro Statistics, 3rd Edition. You can read this by typing
## vignette('os3') or visit www.OpenIntro.org.
##
## The getLabs() function will return a list of the labs available.
##
## The demo(package='DATA606') will list the demos that are available.

##
## Attaching package: 'DATA606'

## The following objects are masked from 'package:openintro':
##
##     calc_streak, present, qqnormsim

## The following object is masked from 'package:utils':
##
##     demo
```

```
library(ggplot2)
library(ggVennDiagram)
library(ggvenn)
```

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

```
library(VennDiagram)
```

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

```
library(knitr)
```

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?

Answer: The probability of rolling a sum of 2 is 0 since it is impossible to roll less than the sum of 2.

(b) getting a sum of 5?

Answer: The set of events to roll a 5 using two dice are {1, 4} or {2,3}. These events are independent of each other (disjoint) so the Additions Rule applies. P(A or B) = P(A) + P(B) = 1/36 + 1/36 = 2/36 = 1/18.

(c) getting a sum of 12?

Answer: Only combination for a sum of 12 is 6 and 6 thus the probability of rolling a sum of 12 is 1/6 x 1/6 = 1/36.

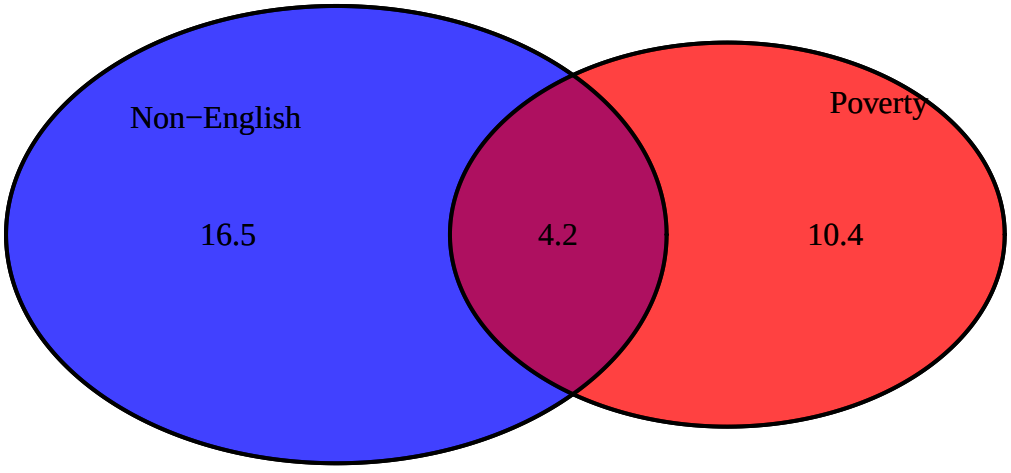
Poverty and language. (3.8, p. 93) The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 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?

Answer: The events are not disjoint (mutually exclusive) because both living below the poverty line and speaking a foreign language can both happen.

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

```
grid.newpage()
venn.plot <- draw.pairwise.venn(
    area1 = 207/10,
    area2 = 146/10,
    cross.area = 42/10,
    category = c("Non-English", "Poverty"),
    fill = c("blue", "red"),
    cex = 1,
    cat.cex = 1,
    cat.pos = c(285, 55),
    cat.dist = 0.09,
    cat.just = list(c(-1.5, -1), c(2.5, 2.25)),
    ext.pos = 30,
    ext.dist = -0.05,
    ext.length = 0.85,
    ext.line.lwd = 2,
    ext.line.lty = "dashed"
    )
grid.draw(venn.plot)
```



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

Answer: 10.4 % of Americans live below the poverty line and only speak English.

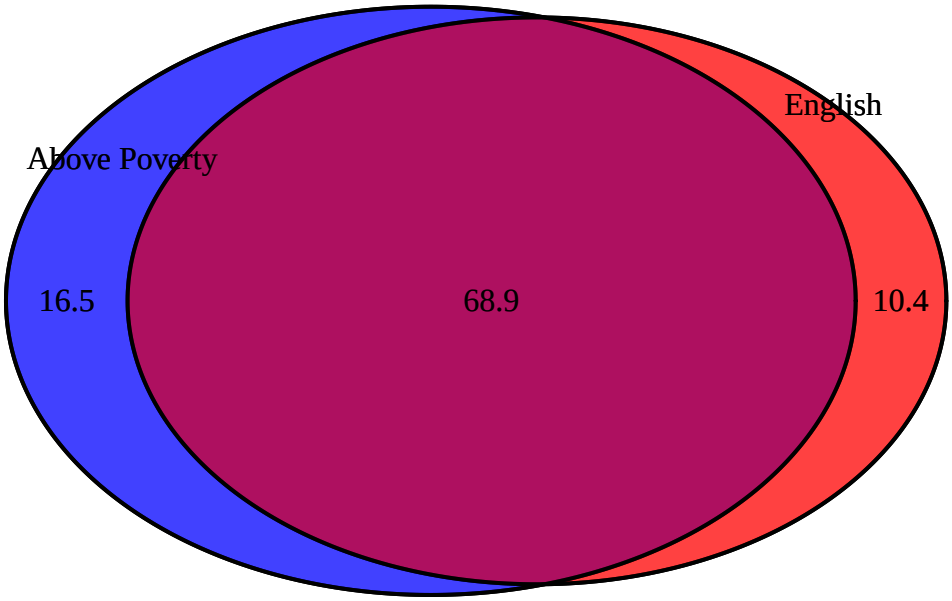
(d) What percent of Americans live below the poverty line or speak a foreign language at home?

Answer: 35.3% (14.6% + 20.7%) of Americans live below the poverty line or speak a foreign language.

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

Answer: 68.9% of Americans who live above the poverty line only speak English.

```
grid.newpage()
venn.plot <- draw.pairwise.venn(85.4, 79.3, 68.9, c("Above Poverty", "English"),    fill = c("blue", "r
  cex = 1,
    cat.cex = 1,
    cat.pos = c(285, 55),
    cat.dist = 0.09,
    cat.just = list(c(-1, -1), c(2.5, 2.5)),
    ext.pos = 30,
    ext.dist = -0.05,
    ext.length = 0.85,
    ext.line.lwd = 2,
    ext.line.lty = "dashed"
    )
grid.draw(venn.plot)
```



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

Answer: The answer here again is no because the events are not disjoint (mutually exclusive) because living below the poverty line and speaking a foreign language can both happen.

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?

Answer: 55.8% (114/204) chance that the male respondent has blue eyes or 52.9% (108/204) chance that 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?

Answer: 78/204 = 38.2% chance that the male repsondent and his partner has blue eyes.

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

Answer: 19/204 = 9.3% chance that a 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?

Answer: 11/204 = 5.4% chance a maler with green eyes has a partner with blue eyes.

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

Answer: It seems that the eye color of male respondents and their partners are not independent because in every case the males picked females with the same eye color the majority of the time.

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.

Answer: Probability of picking a hardcover book first, 28/95. Without replacement (not putting the book back), the probability of picking a paperback book next is 67/94 therefore .295 x .713 = .210 or 21%

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

Answer: Probability of drawing fiction book first is 72/95 = .758. If you pick a hardcover fiction book first then the probability of picking a hardcover book second is 27/94 = .287. If you pick a paperback fiction book first then the probability of picking a hardcover second is 28/94 = .298. Scenario 1 probability = .758 x .287 = .218 or 21.8%. Scenario 2 probability = .758 x .298 = .226 or 22.6%.

(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.

Answer: Probability of drawing fiction book first is 72/95 = .758. If you pick a hardcover fiction book first then the probability of picking a hardcover book second is 28/95 = .295. If you pick a paperback fiction book first then the probability of picking a hardcover second is 28/95 = .295. Scenario 1 probability = .758 x .295 = .224 or 22.4%. Scenario 2 probability = .758 x .295 = .224 or 22.4%.

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

Answer: The final answwers in parts (b) and (c) are very similar in that you have two scenarios in each case but the difference in each although small is that in part (b) you must reduce the inventory by 1 since you do not return the books selected to inventory for the second pick. In part (c), the first pick is returned to inventory.

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.

	Luggage			
I	0	1	2	Total
xi	0	25	35	
P(X=xi)	0.54	0.34	0.12	
xi*P(X=xi)	0	8.50	4.20	12.70
xi - mean	-12.70	12.30	22.30	
(xi-mean)2	161.29	151.29	497.29	
(xi-mean)2*P(X=xi)	87.10	51.44	59.68	198.22

(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.

Answer: Standard Deviation = $14.08. Based on 120 passengers on the flight, for the first bag the revenue will be 120 x .34 x 25 = $1020. For the second bag, the revenue will be 120 x .12 x 35 = $504 for a total of $1524.

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?

Answer: 59,973,543 / 96,420,486 = .622 probability that a randomly chosen resident makes less than \$50,000 a 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.

Answer:

(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.

Answer: