HW_3_DATA606

Salma Elshahawy

9/12/2019

Contents

HomeWork_3 Probability **1**

Dice rolls: 1

Poverty and language 1

Assortative mating. 4

Books on a bookshelf 5

Baggage fees 6

Income and gender. 7

HomeWork_3 Probability

Dice rolls:

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

(a) getting a sum of 1?

0

(b) getting a sum of 5?

4/36 (2,3), (3,2), (1,4), (4,1)

(c) getting a sum of 12?

1/36 (6,6)

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 they are not - they are not disjoint, where 4.2% fall into both categoreis according to The American Community Survey.

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

```
library(VennDiagram)
```

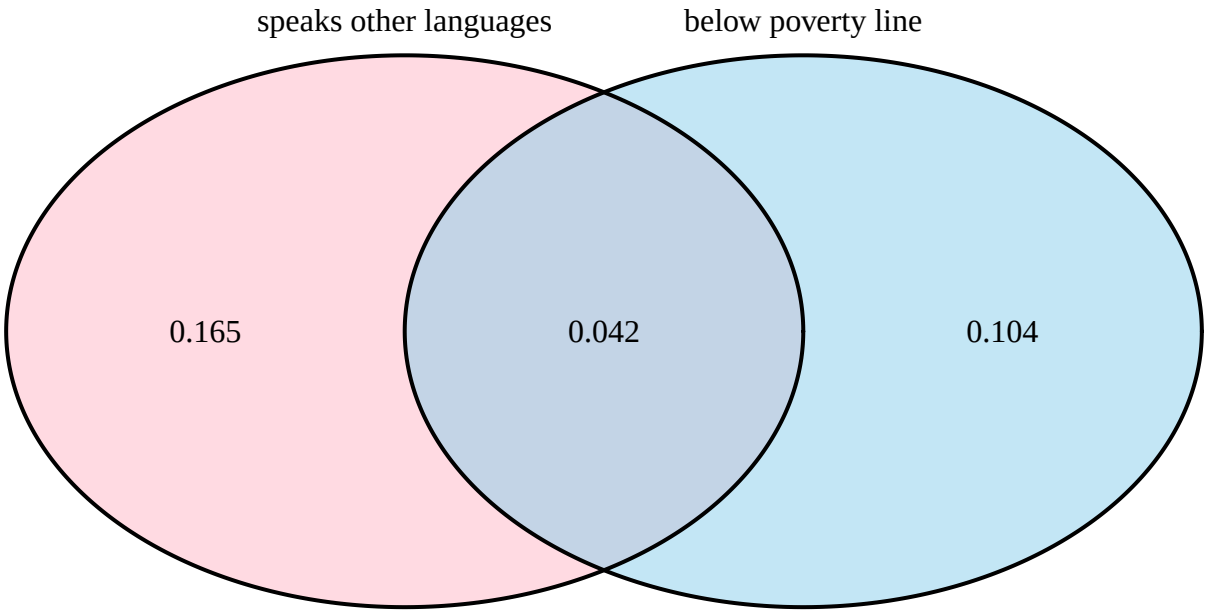
```
## Loading required package: grid
```

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

```
#load the data
survey_df <- c("below poverty line" = 0.146, "speaks other languages" = 0.207, "both" = 0.042)
survey_df
```

```
##     below poverty line speaks other languages                    both
##                  0.146                  0.207                  0.042
```

```
# drawing the diagram
grid.newpage()
draw.pairwise.venn(area1 = 0.146, area2 = 0.207, cross.area = 0.042, category = c("below poverty line",
    fill = c("skyblue", "pink1"),
    alpha = rep(0.5, 2),
    cat.pos = c(0, 0),
    cat.dist = rep(0.025, 2),
    scaled = FALSE)
```




```
## (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?

10.4%

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

```
16.5 + 4.2 + 10.4
```

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

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

```
100 - 31.1
```

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

		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

Figure 1: table

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

```
# If A and B represent events from two different and independent processes, then the probability that b

# P (A and B) = P (A) × P (B)
# P(below poverty line & speaks foriegn language)
0.042 == 0.146 * 0.207
```

```
## [1] FALSE
```

since the two side have different values, so they are not independent.

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?

```
# P(male with blue eyes | female with blue eyes)
# (total male + total female) - common criteria / total population

((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?

```
# P(male with blue eyes & female with blue eyes)
 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?

```
19/54
```

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

```
11/36
```

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

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

```
114/204 == 108/204
```

```
## [1] FALSE
```

No, they are not independent because the two process are not independent from each other. and because both sides of the multiplication rule are not equal to each other.

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.

```
(28/95) * (59/94)
```

```
## [1] 0.1849944
```

		Format		
		Hardcover	Paperback	Total
Type	Fiction	13	59	72
	Nonfiction	15	8	23
	Total	28	67	95

Figure 2: table2

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

```
(72/95) * (28/94)
```

```
## [1] 0.2257559
```

(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.

```
(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.

They are simillar to each other because the possible events are considerable large so the outcome will not be affected by much.

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.

```
mean <- 0*.54 + 25*.34 + (25+35)*.12 #average revenue

ave_rev <- mean
ave_rev
```

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

```
sd <- sqrt(0^2*.54 + 25^2*.34 + 60^2*.12) #standard deviation
sd
```

```
## [1] 25.387
```

(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.

```
expected_rev <- 120 * mean
expected_rev
```

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

```
standard_deviation <- sqrt(120 * sd^2)
standard_deviation
```

```
## [1] 278.1007
```

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.

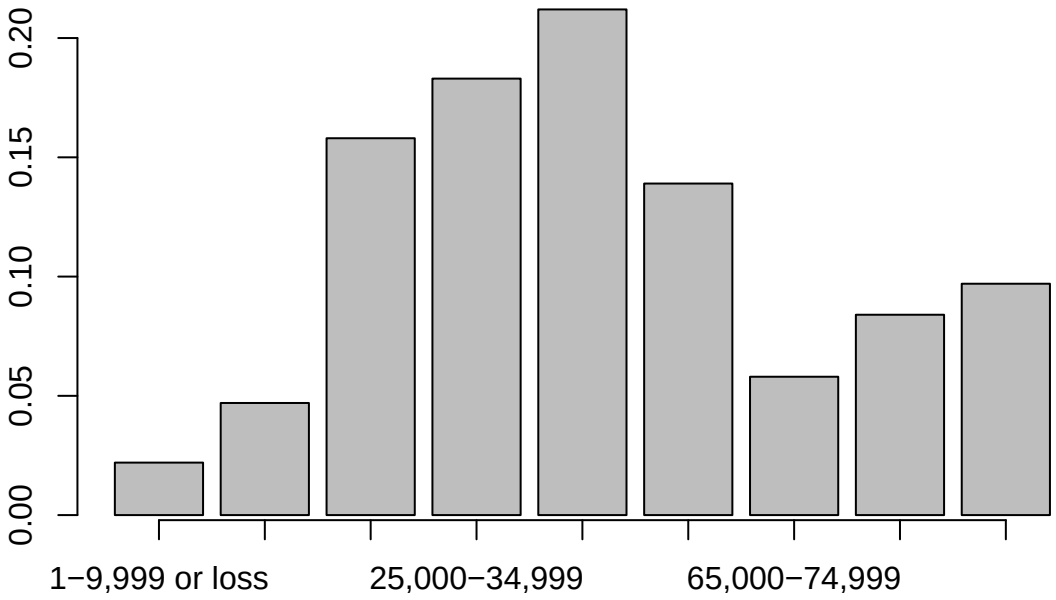
(a) Describe the distribution of total personal income.

```
gender_income <- data.frame(income = c("1-9,999 or loss", "10,000-14,999", "15,000-24,999", "25,000-34,9

counts <- barplot(gender_income$total)
axis(1, at = counts, labels = gender_income$income)
```

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%

Figure 3: table3



The distribution is unimodal and symmetric centered at 35,000-49,999 (Peak point).

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

```
less_50 <- .02 + .047 + .158 + .183 + .212
less_50
```

```
## [1] 0.62
```

(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.

```
less_50_f <- less_50 * 0.41
less_50_f
```

```
## [1] 0.2542
```

Assuming 59% males and 41% females distributed evenly across all income levels.

(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.

71.8% is a big difference from 25.42% I obtained from part (c), therefore the assumption that male and female proportions are distributed evenly across all income levels is not valid.