Chapter 3 - Probability

Leticia Salazar

2021-09-17

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?

- **The probability of getting the sum of 1 is 0 since the least sum we can get is a 2.**

(b) getting a sum of 5?

- **The probability of getting the sum of 5 is 4/36 or 11.11%. There is 4 probabilities being 1+4, 2+3, 3+2, and 4+1.**

(c) getting a sum of 12?

- **The probability of getting the sum of 12 is 1/36 or 2.78%. There is 1 probability being 6+6.**

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, there are 4.2% that fall into both categories therefore it is not disjointed.**

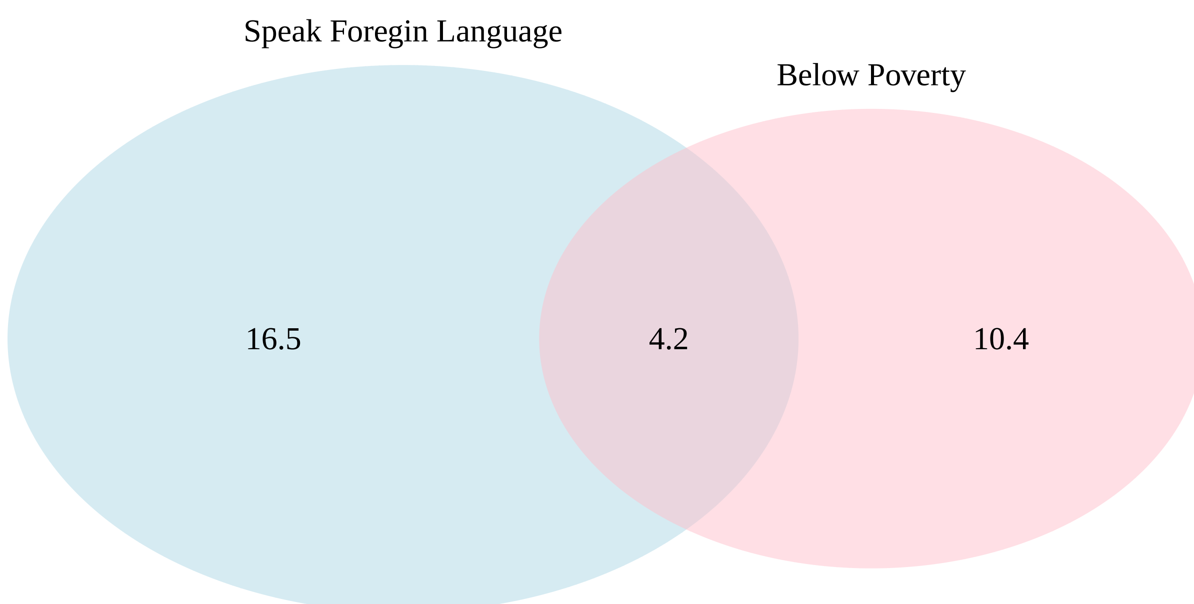
(b) Draw a Venn diagram summarizing the variables and their associated probabilities.

```
library("VennDiagram")
```

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

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

```
grid.newpage()
draw.pairwise.venn(20.7, 14.6, 4.2, category = c("Speak Foregin Language", "Below Poverty"), lty = rep(
    2), fill = c("light blue", "pink"), alpha = rep(0.5, 2), cat.pos = c(0,
    0), cat.dist = rep(0.025, 2))
```




```
## (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(Below Poverty) - P(Below Poverty and Speak Foreign Language) = P(Below Poverty and Speak English).**

```r
14.6 - 4.2
```

```
## [1] 10.4
```

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

- **P(Below Poverty) + P(Speak Foreign Language) - P(Below Poverty and Speak Foreign Language) = P(Below Poverty or Speak Foreign Language).**

```r
14.6 + 20.7 - 4.2
```

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

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

- **P(Total % of Population) - (P Below Poverty and Speak English) = P(Above Poverty and Speak English).**

```r
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?

- **There doesn't seem evidence where this would be independent from each other.**

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?

- **P(Male Blue Eyes) + (Female Blue Eyes) - P(Both Blue Eyes) = P(Male or Female Blue Eyes) or 70.59%**

```r
(114/204) + (108/204) - (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 Blue Eyes) / P(Total Male Blue Eyes) = P(Female Blue Eyes) or 68.42%**

```r
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?

- **P(Male Brown Eyes) / P(Total Male) = P(Male Brown Eyes and Female Blue Eyes) or 35.19%**

```r
19 / 54
```

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

- **P(Male Green Eyes) / P(Total Male) = P(Male Green Eyes and Female Blue Eyes) or 30.56%**

```r
11 / 36
```

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

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

- **When looking at male and female eye colors individually the probability of a random male to have blue eyes is 55.89% and for female it is 52.94%. Due to the high possibilities, they depend on one another.**

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.

- **The probability of drawing a hardcover book first then a paperback fiction without replacement is 18.50%.**

```
#P(Hardcover First) x P(Paperback Fiction)
(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.

- **The probability of drawing a fiction book first and then a hardcover second without replacement is 22.58%**

```
#P(Fiction) x P(Hardcover)
(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.

- **The probability of drawing a fiction book first and then a hardcover second after placing the first book first is 22.33%**

```
#P(Fiction First) x P(Hardcover Second)
(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 very similar because there are only fiction books taken into account with large quantities so the probabilities won't be affected as much.**

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.

```
#Build data frame
Number_of_Bags <- c(0, 1, 2)
Cost_per_Bags <- c(0, 25, 60)
Percentage <- c(.54, .34, .12)

Fees <- data.frame(Number_of_Bags, Cost_per_Bags, Percentage)
Fees
```

```
##   Number_of_Bags Cost_per_Bags Percentage
## 1              0             0       0.54
## 2              1            25       0.34
## 3              2            60       0.12
```

```
#Calculate average revenue per passenger
Avg <- sum(Cost_per_Bags * Percentage)
Avg
```

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

- **The average revenue per passenger is \$15.70**

```
#Calculate variance
var <- (0 - 15.7)^2 * .54 + (25 - 15.7)^2 * .34 + (60 - 15.7)^2 * .12
var
```

```
## [1] 398.01
```

```
#Calculate standard deviation
sd <- sqrt(var)
sd
```

```
## [1] 19.95019
```

- **The standard deviation is 19.95**

(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.

```
#Find revenue on 120 flights
rev <- (Avg * 120)
rev
```

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

- **Expect revenue for 120 passangers is \$1,884**

```r
#Find standard deviation for 120 flights
sd2 <- sqrt(var * 120)
sd2
```

```
## [1] 218.5434
```

- **Standard deviation is 218.54 from the revenue for 120 passengers**

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.

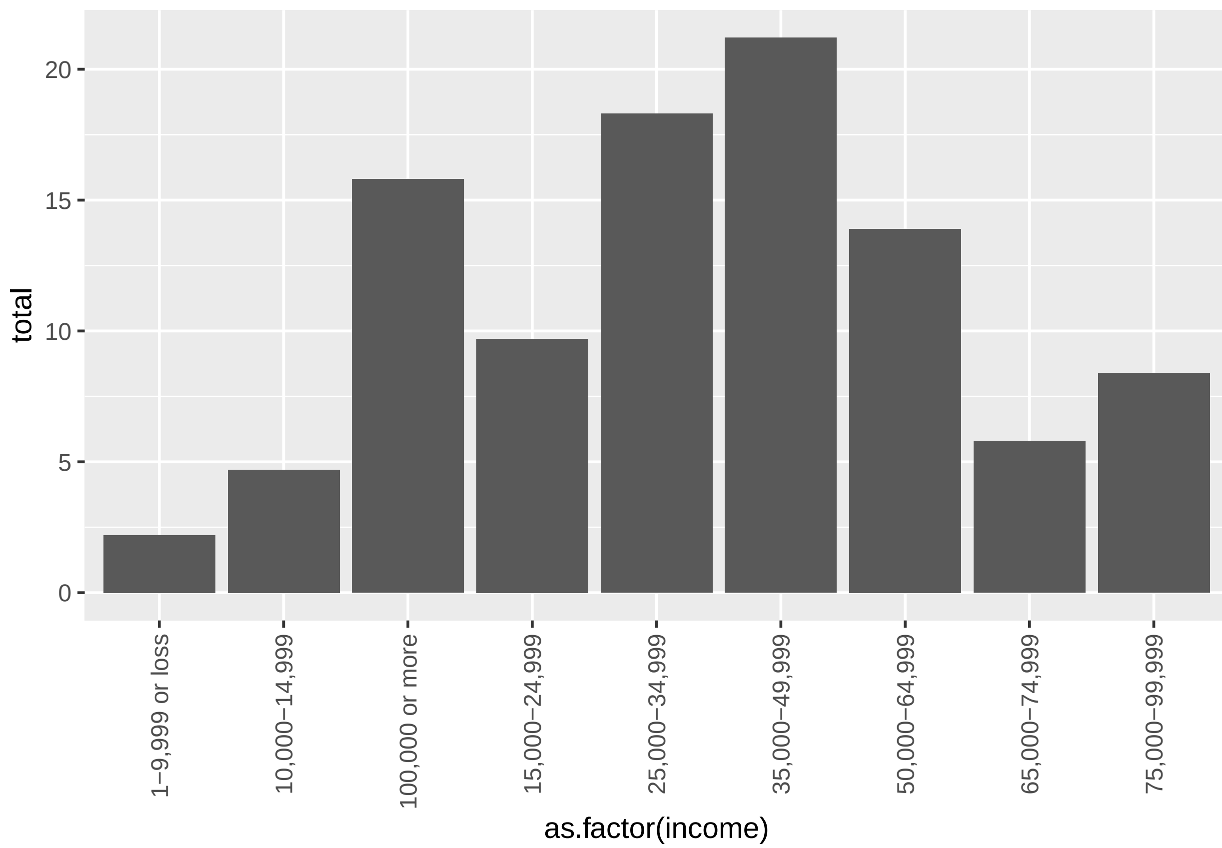
Income	*Total*
$1 to $9,999 or loss	2.2%
$10,000 to $14,999	4.7%
$15,000 to $24,999	15.8%
$25,000 to $34,999	18.3%
$35,000 to $49,999	21.2%
$50,000 to $64,999	13.9%
$65,000 to $74,999	5.8%
$75,000 to $99,999	8.4%
$100,000 or more	9.7%

(a) Describe the distribution of total personal income.

```
#Load Library
library(ggplot2)
#Create data frame and bar plot

data<- data.frame(
  income = c("1-9,999 or loss", "10,000-14,999", "15,000-24,999", "25,000-34,999", "35,000-49,999", "50
  total = c(2.2, 4.7, 9.7, 18.3, 21.2, 13.9, 5.8, 8.4, 15.8)
  ) #There is a mishap where the 100,000+ value was reordered so I reordered the "total" values as well

ggplot(data, aes(x = as.factor(income), y = total)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(angle=90, vjust=.5, hjust=1))
```



- **Based on the barplot above, the distribution seems to be symmetrically multimodal.**

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

- **The probability of randomly chosing US resident that makes less than $50,000 per year is 62.2%**

```
Probability <- sum(2.2 + 4.7 + 15.8 + 18.3 + 21.2)
Probability
```

```
## [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.

- **The probability of randomly choosing a US resident that makes less than $50,000 per year and is female is 25.50% assuming the female population is evenly distributed.**

```
Probability_Female <- 0.41 * Probability
Probability_Female
```

```
## [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.

- **Basing on the source indication and the assumption I made above, the female population is not evenly distributed throughout the data.**