Homework 3 - Probability

Irene Jacob

2020-09-13

## Dice rolls

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

1. getting a sum of 1?

It is not possible to get a sum of 1 when a pair of dice is rolled. The minimum sum possible is 2.

1. getting a sum of 5?

It is possible to get a sum of 5 when a pair of dice is rolled. The different possible combinations are: 1 and 4, 2 and 3, 3 and 2, 4 and 1 So the probability is (4/36)=(1/9)

1. getting a sum of 12?

It is possible to get a sum of 12. There is only 1 possible combination to get it which is 6 and 6. So the probability is (1/36).

## 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.

1. Are living below the poverty line and speaking a foreign language at home disjoint?

There are 4.2% of people fall under both category so it is not disjoint.

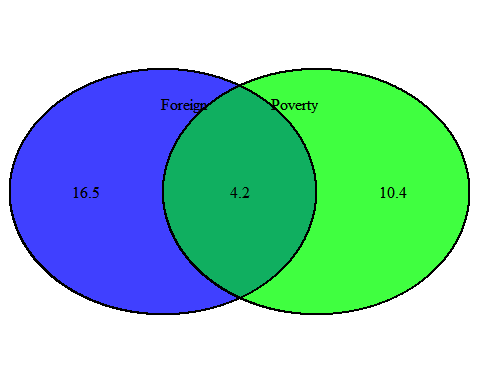
1. Draw a Venn diagram summarizing the variables and their associated probabilities.

library(VennDiagram)

## Loading required package: grid

## Loading required package: futile.logger

library(grid)  
ven\_d <- draw.pairwise.venn(14.6, 20.7, 4.2, c("Foreign", "Poverty"), scale = FALSE, fill = c("green", "blue"))  
grid.draw(ven\_d)



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

From the above venn diagram we can see that 10.4% of the people in the poverty category speak English.

1. What percent of Americans live below the poverty line or speak a foreign language at home?

This would mean 14.6+20.7-4.2=31.1%. This is because below poverty line or speak foreign language means the people on the common region should be ignored.

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

This would mean 100-(14.6+20.7-4.2)=68.9%.

1. Is the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home?

The event that someone lives below the poverty line is not independent of the event that the person speaks a foreign language at home

## 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.

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

The probability is (114+19+11)/204=0.706

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

The probability is (78/114)=0.684

1. 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?

The probability is (19/54)=0.352

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

The probability is (11/36)=0.306

## 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.

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

The probability is ((28/95) \* (59/94))=0.185

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

The probability is ((72/95) \* (28/94))=0.226

1. 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 is ((72/95) \* (28/95))=0.2234

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

The above sample considers a a large set of 95 books so when one book is removed or added it does not make much difference to the final output.

## 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.

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

n\_bags <- c(0, 1, 2)  
fee <- c(0,25,35)  
pass <- c(0.54, 0.34, 0.12)  
baggage <- data.frame(n\_bags,fee,pass)  
print(baggage)

## n\_bags fee pass  
## 1 0 0 0.54  
## 2 1 25 0.34  
## 3 2 35 0.12

The average revenue of passenger is

avg <- sum((baggage$pass\*baggage$fee))/sum(baggage$pass)  
print(avg)

## [1] 12.7

The standard deviation is

sd <- sqrt(0.54\*(0-avg)^2 + 0.34\*(25-avg)^2 + 0.12\*(35-avg)^2)  
print(sd)

## [1] 14.07871

1. 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.

Based on the input given for the case of 120 passengers lets assume that 65 have 0 checked in bags, 41 have 1 checked in bag and 14 have 2 checked in bags. The revenue is:

revenue\_120 <- (65\*0 + 41\*25 + 14 \* 35)  
print(revenue\_120)

## [1] 1515

The standard deviation is:

avg\_120 <- revenue\_120/120  
sd\_120 <- sqrt(0.54\*(0-avg\_120)^2 + 0.34\*(25-avg\_120)^2 + 0.12\*(35-avg\_120)^2)  
print(sd\_120)

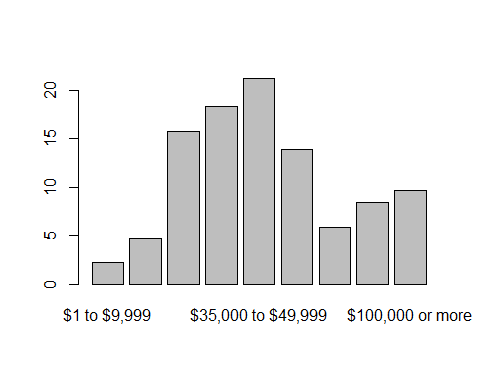
## [1] 14.07891

## 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.

1. Describe the distribution of total personal income.

dist <- data.frame(  
i <- c("$1 to $9,999","$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"),  
t <- c(2.2,4.7,15.8,18.3,21.2,13.9,5.8,8.4,9.7))  
  
barplot(dist$t, names.arg=i)



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

The probability that a randomly chosen US resident makes less than $50,000 per year is (2.2 + 4.7 + 15.8 + 18.3 + 21.2)= 62.2 %

1. 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 that a randomly chosen US resident makes less than $50,000 per year and is female is (0.622\*0.41)=0.255=25.5%

Note: The assumption made was that the income and gender are independent of each other.

1. 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 that income and gender and independent but from the data saying that 71.8% are females it is clear that these two are actually dependent.