

606-01__Week2__HW__Probability

Shyam BV

September 10, 2016

2.6 Dice rolls. If you roll a pair of fair dice, what is the probability of

```
dice_roll_probability <- c(1,2,3,4,5,6,5,4,3,2,1)/36
dice_roll_probability
```

```
## [1] 0.02777778 0.05555556 0.08333333 0.11111111 0.13888889 0.16666667
## [7] 0.13888889 0.11111111 0.08333333 0.05555556 0.02777778
```

(a) **getting a sum of 1?** Sum of 1 cannot happen in two dice. Minimum is 2. So the probability is 0

(b) **getting a sum of 5?** From the above probability chart, sum of 5 is 0.11111111

(c) **getting a sum of 12?** From the above probability chart, sum of 12 is 0.02777778

2.8 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 disjoint.

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

Diagram is shown below

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

$P(\text{Americans poverty} \mid \text{Speak english}) = P(A \text{ and } S)/P(S)$

```
.042/.165
```

```
## [1] 0.2545455
```

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

$P(A \text{ or } S) = P(A) + P(S) - p(A \text{ and } S)$

```
(.146+.207) - .042
```

```
## [1] 0.311
```

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

```
# Above poverty line =
(100 - (10.4+4.2+16.5)) /100
```

```
## [1] 0.689
```

(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.

$P(\text{poverty americans and foreign speaking}) \neq P(\text{poverty americans}) \times P(\text{foreign speaking})$

```
# 0.042 != .165*.104
```

2.20 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.

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

```
assortive_mating <- read.csv("https://raw.githubusercontent.com/jbryer/DATA606Fall2016/master/Data/Dataa.csv")
```

```
table(assortive_mating)
```

```
##           partner_female
## self_male blue brown green
##    blue    78    23    13
##   brown    19    23    12
##   green    11     9    16
```

```
#  $P(M) + P(F) - P(M \text{ and } F) =$ 
```

```
((sum(assortive_mating$self_male == "blue")/nrow(assortive_mating)) +  
 (sum(assortive_mating$partner_female == "blue")/nrow(assortive_mating)))- (sum(assortive_mating$self_ma
```

```
## [1] 0.7058824
```

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

```
(sum(assortive_mating$self_male == "blue" & assortive_mating$partner_female=="blue")/nrow(assortive_mating))
```

```
## [1] 0.3823529
```

```
#  $P(MB \mid FB) =$ 
```

```
(sum(assortive_mating$self_male == "blue" & assortive_mating$partner_female=="blue")/nrow(assortive_mating))
```

```
## [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(MB | FB) =
```

```
(sum(assortive_mating$self_male == "brown" & assortive_mating$partner_female == "blue")/nrow(assortive_m
```

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

```
#P(MG | FB) =
```

```
(sum(assortive_mating$self_male == "green" & assortive_mating$partner_female == "blue")/nrow(assortive_m
```

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

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

```
(sum(assortive_mating$self_male == "blue" & assortive_mating$partner_female == "blue")/nrow(assortive_mat
```

```
## [1] 0.3823529
```

```
# !=
```

```
(sum(assortive_mating$self_male == "blue")/nrow(assortive_mating))
```

```
## [1] 0.5588235
```

```
# No. They are not independent. Probability of blue male and female eyes are not equal to female blue e
```

2.30 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.

```
books <- read.csv("https://raw.githubusercontent.com/jbryer/DATA606Fall2016/master/Data/Data%20from%20o  
table(books)
```

```
##          format  
## type      hardcover paperback  
##  fiction         13         59  
## nonfiction         15          8
```

```
sum(books$format == "hardcover")
```

```
## [1] 28
```

```
# P(H) * P(P)
```

```
(sum(books$format == "hardcover"))/sum(table(books)) *  
(sum(books$format == "paperback"))/(sum(table(books))-1)
```

```
## [1] 0.2100784
```

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

```
(sum(books$type=="fiction"))/sum(table(books)) *
(sum(books$format=="hardcover"))/(sum(table(books))-1)
```

```
## [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.

```
(sum(books$type=="fiction"))/sum(table(books)) *
(sum(books$format=="hardcover"))/sum(table(books))
```

```
## [1] 0.2233795
```

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

It is similar because there is only a fraction of difference between both. But when we take the actual number, you can see the difference.

```
(sum(books$type=="fiction"))/sum(table(books)) *
(sum(books$format=="hardcover"))/sum(table(books)) - (sum(books$type=="fiction"))/sum(table(books)) *
(sum(books$format=="hardcover"))/(sum(table(books))-1)
```

```
## [1] -0.002376378
```

2.38 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.

```
library("htmlTable")
fees <- c(0,25,60)
passengers <- c(54,34,12)
baggage <- data.frame(fees,passengers)

## Average revenue
average_revenue <- (sum((baggage$fees*baggage$passengers)))/sum(baggage$passengers)

## 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.

```
##Total airline expected revenue for 120 passengers is 1884
```

```
## (120*.34)*25 + (120*.12)*35 +(120*.12)*25
```

```
## With the standard deviation of 19.95
```

2.44 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.

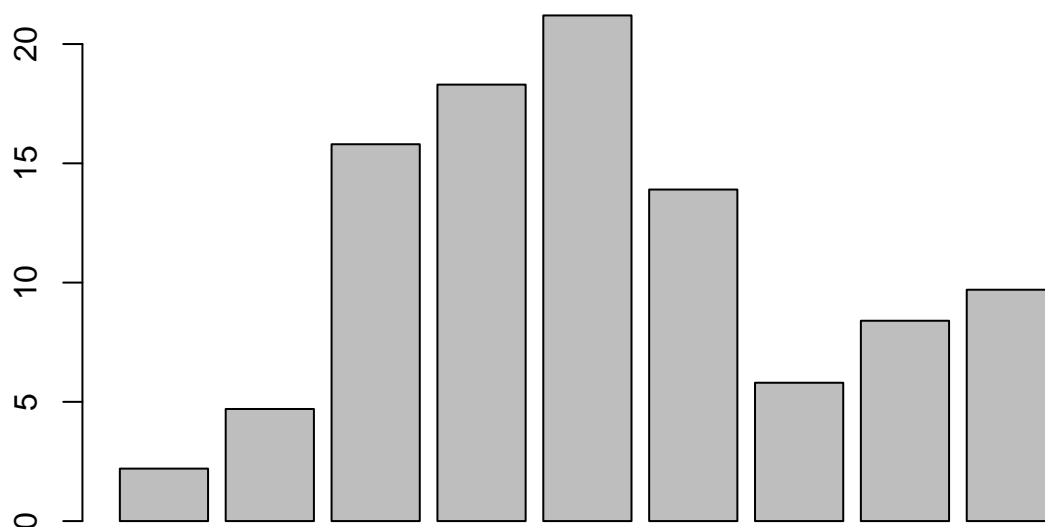
(a) Describe the distribution of total personal income.

```
income <- 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")
total <- c(2.2, 4.7, 15.8, 18.3, 21.2, 13.9, 5.8, 8.4, 9.7)

frequency_table <- data.frame(income, total)
frequency_table
```

```
##           income total
## 1      $1 to $9,999  2.2
## 2 $10,000 to $14,999  4.7
## 3 $15,000 to $24,999 15.8
## 4 $25,000 to $34,999 18.3
## 5 $35,000 to $49,999 21.2
## 6 $50,000 to $64,999 13.9
## 7 $65,000 to $74,999  5.8
## 8 $75,000 to $99,999  8.4
## 9  $100,000 or more  9.7
```

```
barplot(frequency_table$total)
```



This distribution has dual peak. The mean is centered around 25K to 49K.

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

Probability that a randomly chosen US resident makes less than \$50,000 per year is 0.622

(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.

Randomly chosen US resident makes less than \$50K are 0.25502

(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.

If 71.8% females make less than 50K then 0.29438 is the total women make less than \$50K. It is not equal to 0.25502. So they are dependent.